

Akshi Kumar  
Abhishek Swaroop  
Pancham Shukla *Editors*

# Proceedings of Fourth International Conference on Computing and Communication Networks

ICCCN 2024, Volume 4

# **Lecture Notes in Networks and Systems**

**Volume 1292**

## **Series Editor**

Janusz Kacprzyk , Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

## **Advisory Editors**

Fernando Gomide, Department of Computer Engineering and Automation—DCA, School of Electrical and Computer Engineering—FEEC, University of Campinas—UNICAMP, São Paulo, Brazil

Okyay Kaynak, Department of Electrical and Electronic Engineering, Bogazici University, Istanbul, Türkiye

Derong Liu, Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, USA

Institute of Automation, Chinese Academy of Sciences, Beijing, China

Witold Pedrycz, Department of Electrical and Computer Engineering, University of Alberta, Alberta, Canada

Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Marios M. Polycarpou, Department of Electrical and Computer Engineering, KIOS Research Center for Intelligent Systems and Networks, University of Cyprus, Nicosia, Cyprus

Imre J. Rudas, Óbuda University, Budapest, Hungary

Jun Wang, Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong

The series “Lecture Notes in Networks and Systems” publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, EI Compendex, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose ([aninda.bose@springer.com](mailto:aninda.bose@springer.com)).

Akshi Kumar · Abhishek Swaroop ·  
Pancham Shukla  
Editors

# Proceedings of Fourth International Conference on Computing and Communication Networks

ICCCN 2024, Volume 4



Springer

*Editors*

Akshi Kumar

Department of Computing  
Goldsmiths University of London  
London, UK

Abhishek Swaroop

Bhagwan Parshuram Institute  
of Technology  
New Delhi, Delhi, India

Pancham Shukla

Department of Computing  
Faculty of Engineering  
Imperial College London  
London, UK

ISSN 2367-3370

ISSN 2367-3389 (electronic)

Lecture Notes in Networks and Systems

ISBN 978-981-96-3249-7

ISBN 978-981-96-3250-3 (eBook)

<https://doi.org/10.1007/978-981-96-3250-3>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2025

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.  
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721,

If disposing of this product, please recycle the paper.

# **ICCCNet-2024 Committees**

## **ICCCNet-2024 Steering Committee**

### **General Chairs**

Prof. Dr. Omer Rana, Cardiff University, UK

Prof. Dr. Yang Xiao, The University of Alabama, USA

### **Honorary Chairs**

Prof. Dr. Vincenzo Piuri, University of Milan, Italy

Prof. Dr. Janusz Kacprzyk, FIEEE, Polish Academy of Sciences, Poland

Prof. Dr. Manu Malek, EiC Computer and Electrical Engineering, Stevens Institute of Technology, USA

Dr. Jon G. Hall, EiC Expert Systems (WILEY), The Open University, UK

### **Conference Chairs**

Dr. Ali Kashif Bashir, Manchester Metropolitan University, UK

Prof. Dr. Darren Dancey, Manchester Metropolitan University, UK

Prof. Abhishek Swaroop, Bhagwan Parshuram Institute of Technology, India

## Technical Program Chairs

Dr. Rajkumar Rathore, Cardiff Metropolitan University, UK  
Fatma Bassyouni, National Research Centre, Cairo, Egypt  
Aditya Khamparia, Babasaheb Bhimrao Ambedkar University, India

## Conveners

Dr. Akshi Kumar, Goldsmiths University of London, UK  
Dr. Pancham Shukla, London Metropolitan University, London, UK  
Dr. Utku Kose, Suleyman Demirel University, Isparta, Turkey  
Dr. Harpreet Singh Arora, Tel-Aviv University, Israel

## Publicity Chairs

Prof. Dr. Valentina Emilia Balas, Aurel Vlaicu University of Arad, Romania  
Prof. Dr. Zdzislaw Polokowski, Jan Wyzykowski University, Polkowice, Poland  
Prof. Dr. Vijay Singh Rathore, IIS (Deemed to be University), Jaipur  
Dr. Hamid Reza Boveiri, Sama College, Islamic Azad University, Shoushtar Branch, Iran  
Dr. Syed Bilal Shah, Darul Hekma University, Saudi Arabia  
Dr. Umesh Gupta, Bennett University, India  
Dr. Bilal Ozturk, Software Engineering Department, Istanbul Aydin University, Istanbul, Turkey

## ICCCNet-2024 Advisory Committee and Technical Program Committee

Prof. Dr. Vincenzo Piuri, University of Milan, Italy  
Prof. Dr. Valentina Emilia Balas, Aurel Vlaicu University of Arad, Romania  
Prof. Dr. Aboul Ella Hassanien, Cairo University, Egypt  
Prof. Dr. Vineet Kansal, ProVC, AKTU; Director, IET, Lucknow, India  
Prof. Dr. João Manuel R. S. Tavares, Universidade do Porto (FEUP), Portugal  
Prof. Dr. Neeraj Kumar, Thapar Institute of Engineering and Technology, Patiala, Punjab, India  
Prof. Dr. Zdzislaw Polkowski, Jan Wyzykowski University, Poland  
Prof. Dr. George A. Tsirhrintzis, University of Piraeus, Greece  
Prof. Dr. Arun Sharma, Indira Gandhi Delhi Technical University for Women, India

- Prof. Dr. Abhishek Swaroop, Bhagwan Parshuram Institute of Technology, India  
Prof. Dr. Giorgos Karagiannidis, Aristotle University of Thessaloniki, Greece  
Prof. Dr. Sheng-Lung Peng, National Dong Hwa University, Taiwan  
Dr. Dijana Oreski, University of Zagreb, Varazdin, Croatia  
Dr. Ahmed Zobaa, Brunel University London, UK  
Prof. Dr. A. K. Singh, National Institute of Technology Kurukshetra, India  
Prof. Dr. Anil Kumar Ahlawat, KIET Group of Institutes, India  
Dr. Jafar A. Alzubi, Al-Balqa Applied University, Salt, Jordan  
Prof. Dr. Alex Norta, Tallinn University of Technology, Estonia  
Dr. Utku Kose, Suleyman Demirel University, Isparta, Turkey  
Prof. Dr. Isabel De la Torre Díez, University of Valladolid, Spain  
Dr. Oana Geman, Universitatea Stefan cel Mare din Suceava, Romania  
Dr. Varun G. Menon, SCMS School of Engineering and Technology, Kochi, India  
Dr. Mohammad Shojafar, University of Surrey, UK  
Dr. Muhammad Habib ur Rehman, King's College London, UK  
Dr. Irfan Mehmood, University of Bradford, UK  
Dr. Muddeswar Iqbal, London South Bank University, UK  
Dr. Shahid Mumtaz, Institute de Telecommunication, Portugal  
Dr. Waleed Ejaz, Lakehead University, Canada  
Dr. Anish Jindal, University of Essex, UK  
Dr. Suresh Chavhan, Vellore Institute of Technology, Vellore, India  
Dr. Gagandeet Singh Aujla, Durham University, UK  
Dr. Sachin Kumar, South Ural State University, Chelyabinsk, Russian Federation  
Dr. Prayag Tiwari, Aalto University, Finland  
Dr. Pradeep Malik, KIIT University, India  
Dr. Akash Kumar Bhoi, Sikkim Manipal Institute of Technology, India  
Dr. Hamid Reza Boveiri, Sama College, IAU, Shoushtar Branch, Shoushtar, Iran  
Dr. Sahil Garg, École de technologie supérieure, Université du Québec, Montreal, Canada  
Dr. Gulshan Shrivastava, Sharda University, Greater Noida, India  
Dr. Gabriella Casalino, University of Bari, Italy  
Dr. Aditya Khamparia, Babasaheb Bhimrao Ambedkar University, India  
Dr. Amit Kumar Jaiswal, University of Leeds, UK  
Dr. Qianqian Xie, University of Manchester, Manchester, UK  
Dr. Yousaf Bin Zikria, Yeungnam University, South Korea  
Dr. Francesco Piccialli, University of Naples Federico II, Italy  
Dr. Ashiq Anjum, University of Leicester, UK  
Prof. Dr. Nuno M. Garcia, University of Beira Interior, Covilhã, Portugal  
Prof. Dr. Kashif Saleem, Universiti Teknologi Malaysia, Riyadh, Saudi Arabia  
Dr. Le Hoang Son, University of Dannang, Vietnam  
Dr. Jaafar Alghazo, Virginia Military Institute, Lexington, VA, USA  
Dr. Jalil Piran, Sejong University, South Korea  
Dr. Hari Mohan Pandey, Edge Hill University, UK  
Prof. Dr. P. Sanjeevikumar, Aarhus University, Herning, Denmark  
Dr. Kemal Polat, Abant Izzet Baysal University, Turkey

- Dr. Juhriyansyah Dalle, Universitas Lambung Mangkurat, Indonesia  
Dr. Ahmed Elngar, Beni-Suef University, Egypt  
Dr. Prajjoy Podder, Institute of Information and Communication Technology, BUET, Dhaka, Bangladesh  
Dr. M. Rubaiyat Hossain Mondal, Institute of Information and Communication Technology, BUET, Dhaka, Bangladesh  
Dr. Sarada Prasad Gochhayat, Old Dominion University, USA  
Dr. Daniel Nogueira, University of Minho, Brazil  
Dr. Khan Muhammad, Sejong University, South Korea  
Dr. Yenumula B. Reddy, Grambling State University, Louisiana, USA  
Dr. Chandran Venkatesan, KPR Institute of Engineering and Technology, India  
Dr. Alireza Jolfaei, Macquarie University, Australia  
Dr. Souvik Ganguli, Thapar Institute of Engineering and Technology, India  
Dr. Flah Aymen, National School of Engineering of Gabes, Tunisia  
Prof. Dr. Placido Rogerio Pinheiro, University of Fortaleza, Brazil  
Dr. Daniela Clara Moraru, University of Luxembourg, Luxembourg  
Dr. Gautam Srivastava, Brandon University, Canada  
Dr. Vassilis C. Gerogiannis, University of Thessaly, Greece  
Prof. Dr. B. S. Manoj, Indian Institute of Space Science and Technology, India  
Dr. M<sup>a</sup> Luz Castro Pena, Universidade da Coruña, Spain  
Dr. Ilya Levin, Tel Aviv University, Israel  
Dr. Muhibul Haque Bhuyan, Southeast University, Bangladesh  
Prof. Dr. Med Salim Bouhlel, Lab SETIT; Sfax University, Tunisia  
Dr. Mamoun Alazab, Charles Darwin University, Australia  
Dr. Lalit Garg, University of Malta, Msida, Malta  
Dr. Arij Naser Abougreen, University of Tripoli, Libya  
Dr. Sherif Mohamed Ismail, Egyptian German Academy  
Prof. Dr. Vijay Singh Rathore, IIS Deemed to be University, India  
Dr. Aslanbek Naziev, Ryazan State University named after S. A. Esenin, Russia  
Dr. Mwaffaq Otoom, Yarmouk University, Jordan  
Dr. Ahmed A. Ewees, Damietta University, Egypt  
Dr. Iwan Adhicandra, University of Sydney, Australia  
Prof. Dr. Meng Li, Hefei University of Technology, China  
Dr. Korhan Cengiz, Trakya University, Turkey  
Dr. Muhammad Bilal, Hankuk University of Foreign Studies, South Korea  
Dr. RR Venkatesha Prasad, TU Delft, The Netherlands  
Dr. Özge Korkmaz, Malatya Turgut Özal University, Turkey  
Dr. Alexander Fedorov, Rostov State University of Economics, Russia  
Prof. Dr. Alfredo Grieco, Politecnico di Bari, Italy  
Prof. Dr. Quoc-Viet Pham, Pusan National University, South Korea  
Dr. Enkeleda Lula, University Haxhi Zeka, Peja, Kosovo  
Dr. Fides del Castillo, De La Salle University, Philippines  
Dr. Houda Chihi, Innov'COM Lab of Sup'COM, Tunisia  
Prof. Dr. Tu Nguyen, Kennesaw State University, Kennesaw, USA  
Prof. Dr. Christos Douligeris, University of Piraeus, Greece

Dr. Surbhi Bhatia, King Faisal University, Saudi Arabia

Dr. Feras M. Awaysheh, Tartu University, Delta Research Center, Estonia

Dr. Assunta Di Vaio, University of Naples “Parthenope” (Italy)

Dr. Mehdi Gheisari, Southern University of Science and Technology, Shenzhen, Guangdong Province, P. R. China

Dr. Suleyman Eken, Kocaeli University, Turkey

# Preface

We hereby are delighted to announce that Manchester Metropolitan University, Manchester, UK, has hosted the eagerly awaited and much coveted International Conference on Computing and Communication Networks (ICCCNet-2024) in hybrid mode during October 17–18, 2024. The conference has attracted many high-quality submissions and stimulates the cutting-edge research discussions among many academic pioneering researchers, scientists, industrial engineers, and students with the reception of papers including more than 4800 authors from different parts of the world. The committee of professionals dedicated toward the conference is striving to achieve a high-quality technical program with tracks on Networks and Computing Technologies, Advances in Artificial Intelligence and Machine Learning, Security and Privacy, Emerging Topics in 5G/6G Communication Systems, Cyber physical Systems, Emerging Trends in Data Analytics, Cyber Security for Industry 4.0 and Smart and Sustainable Environmental Systems. All the tracks chosen in the conference are interrelated and are very famous among the present-day research community. Therefore, a lot of research is happening in the above-mentioned tracks and their related sub-areas. More than 1150 full-length papers have been received, among which the contributions are focused on theoretical, computer simulation-based research and laboratory-scale experiments. Among these manuscripts, 304 papers have been included in the Springer proceedings after a thorough two-stage review and editing process. All the manuscripts submitted to the ICCCNet-2024 were peer-reviewed by at least two independent reviewers, who were provided with a detailed review pro forma. The comments from the reviewers were communicated to the authors, who incorporated the suggestions in their revised manuscripts. The recommendations from two reviewers were taken into consideration while selecting a manuscript for inclusion in the proceedings. The exhaustiveness of the review process is evident, given the large number of articles received addressing a wide range of research areas. The stringent review process ensured that each published manuscript met the rigorous academic and scientific standards. It is an exalting experience to finally see that these elite contributions materialize into six volumes as ICCCNet-2024 proceedings by Springer entitled “International Conference on Computing and Communication Networks”.

ICCCNet-2024 invited two keynote speakers, who are eminent researchers in the field of computer science and engineering, from different parts of the world. In addition to the plenary sessions on each day of the conference, twelve (four in physical mode and twenty six in online mode) concurrent technical sessions are held every day to assure the oral presentation of around 304 accepted papers. Keynote speakers and session chair(s) for each of the concurrent sessions have been leading researchers from the thematic area of the session. A technical exhibition is held during all the 2 days of the conference, which has put on display the latest technologies, expositions, ideas and presentations.

An international conference of such magnitude and release of the ICCCNet-2024 proceedings by Springer has been the remarkable outcome of the untiring efforts of the entire organizing team. The success of an event undoubtedly involves the painstaking efforts of several contributors at different stages, dictated by their devotion and sincerity. Fortunately, since the beginning of its journey, ICCCNet-2024 has received support and contributions from every corner. We thank them all who have wished the best for ICCCNet-2024 and contributed by any means toward its success. The edited proceedings volume by Springer would not have been possible without the perseverance of all the steering, advisory and technical program committee members.

All the contributing authors owe thanks from the organizers of ICCCNet-2024 for their interest and exceptional articles. We would also like to thank the authors of the papers for adhering to the time schedule and for incorporating the review comments. We wish to extend my heartfelt acknowledgment to the authors, peer-reviewers, committee members and production staff whose diligent work put shape to the ICCCNet-2024 proceedings. We especially want to thank our dedicated team of peer-reviewers who volunteered for the arduous and tedious step of quality checking and critique on the submitted manuscripts. The management, faculties, administrative and support staff of the university have always been extending their services whenever needed, for which we remain thankful to them.

Lastly, we would like to thank Springer for accepting our proposal for publishing the ICCCNet-2024 conference proceedings. Help received from Mr. Aninda Bose, the acquisition senior editor, in the process has been very useful.

London, UK  
Delhi, India  
London, UK

Akshi Kumar  
Abhishek Swaroop  
Pancham Shukla  
Organizers and Conveners  
ICCCNet-2024

# Contents

<b>Dynamic Weight-Adjusted Ensemble Loss for Enhanced Medical Image Segmentation .....</b>	1
Mohsin Furkh Dar and Avatharam Ganivada	
<b>Stirring up Biomarker Discovery for Cardiovascular Diseases Diagnosis: The FDR-RFE Pipeline .....</b>	13
Harsh Bhasin, Chandsi, and Kapila Kumar	
<b>Support Vector Machines and Slime Mold Optimization Algorithms for SQL Injection Detection .....</b>	21
Zainab H. Al-Araji and Hasanen Alyasiri	
<b>Sustainable Sensory Based Automated Water Regulation System in Energy Constrained Domain .....</b>	37
Ankit Kumar, Saksham Srivastava, Tiansheng Yang, Lu Wang, and Rajkumar Singh Rathore	
<b>Deep Neural Network with Stochastic LWTA (DNN-S-LWTA) for Adversarial Machine Learning .....</b>	47
Ms. Soumya and Robin Rohit Vincent	
<b>Stock Price Prediction Using Arithmetic Optimizer-Assisted LSTM Model .....</b>	55
P. V. Bhuvaneshwari, Radhakrishnan Vignesh, and H. B. Asif Mohamed	
<b>Studying the Differences in Functioning Among Several Effective Image Steganography Methods .....</b>	65
Raghda Salam Al. Mahdawi, Warqaa Shaher Alazawee, Ali J. Abboud, and Azmi Shawkat Abdulbaqi	
<b>Monitoring and Optimization of Machine Learning Workloads Using Kubernetes .....</b>	89
Ananth Mahesh Kashyap, V. Dinesh Reddy, and Marco Aiello	

<b>CO<sub>2</sub> Emissions of AI Applications: An Investigation on its Measurement .....</b>	99
Pankhuri Verma, V. Dinesh Reddy, and Marco Aiello	
<b>Research on Optimization of Machine Translation Performance Based on Deep Learning Algorithm .....</b>	111
Yan Meng and Jiuquan Zhang	
<b>Using Big Data Analytics and Business Intelligence for Flight Delay Prediction .....</b>	121
Mona Hassan Asiri, Abdullah S. A. L.-Malaise AL-Ghamdi, Ayman G. Fayoumi, and Mahmoud Ragab	
<b>A Location-Specific Mobile Framework for Intelligent Road Traffic Traceability Systems .....</b>	141
Khushi Saxena, Tiansheng Yang, Ruikai Sun, Changgui Lin, Lu Wang, and Rajkumar Singh Rathore	
<b>Toward Efficient Multi-attribute Prediction: Lessons from Political Bias Detection .....</b>	149
Charan Ramtej Kodi, Satya Sai Bharath Vemula, Nikhil Kumar Pulipeta, Varsha Venkata Krishnan, Vishal Rajkumar, and Bharath Goud Musalaya	
<b>Sentiment Analysis and Assessment of Public Opinions Regarding COVID-19 Vaccination via Twitter and Machine Learning Techniques .....</b>	165
Roa'a Mohammedqasem, Hayder Mohammedqasim, Bilal A. Ozturk, Layth Mhmod Farhan, and Abualqasim Khalil Ismael	
<b>Artificial Intelligence Driven Kyphosis Classification .....</b>	177
V. Thamilarasi, R. Harihara Krishnan, V. Vijayalakshmi, J. Mary Catherine, V. Poornima, and S. Pratheepa	
<b>An Enhanced Lightweight Authentication Scheme Based on Three Factors for WSN .....</b>	191
Shilpi Sharma and Bijendra Kumar	
<b>EfficientDet with SAM on NC4K Dataset .....</b>	205
Lavish Kumar and Shweta Meena	
<b>Blockchain for Cloud/Edge/Fog Computing: A Review .....</b>	219
Soukeina Zouaidi	
<b>Secure Data Communication: Implementation and Performance Evaluation of Aggregate Key Encryption .....</b>	231
Nagabhryu Nikitha, Kogatam Vijayasree, Nara Bhavyasree, and D. Radha	

<b>Advancing Pneumonia Diagnosis: Hybrid and Optimal Deep CNN Model for Chest Image Classification .....</b>	245
Gunapati Suresh, T. Ravi, and R. Krishnaprasanna	
<b>Enhancing Intrusion Detection by Integrating Deep Learning and Traditional Machine Learning Techniques .....</b>	261
Noor Yeshfeen and Ritika Kumari	
<b>Strategic Safeguards: Fortifying Sovereign Tender Security with RNNs and Multi-focal Attention .....</b>	273
Rishabh Mohata, Akash Chandrakar, Tiansheng Yang, Rajkumar Singh Rathore, Aaryan Raj, and Hrudaya Kumar Tripathy	
<b>Integrating Machine Learning into Cardiovascular Disease Risk Prediction: A Comprehensive Analysis of Cholesterol, Heart Rate, and Gender Impact on Disease Prevalence .....</b>	285
Abdul Rahim, Amit Chhabra, Manya, Sunil K. Singh, Sudhakar Kumar, Hardik Gupta, and Karan Sharma	
<b>Review of Machine Learning and False Advertising in Live E-commerce: Features, Motivations, and Identification Studies .....</b>	297
Tingsen Gan, Kelang Yang, and Wei Wang	
<b>Optimization of Digital Special Effects Innovation Technology for LCS Algorithm .....</b>	307
Anjia Ma	
<b>Human-Centric Video Analysis in Industrial Environments .....</b>	319
Hayder Mohammedqasim, Roa'a Mohammedqasem, Bilal A. Ozturk, Habib Rahman Hamedy, and Ali bin Asghar	
<b>Survey and Analysis of Communication and Routing Mechanisms for UAVs in Wireless Sensor Networks .....</b>	333
Prajoy Podder and Maciej Zawodniok	
<b>Securing Patient Personal Information Using Multi-Dimensional Anonymization-Based Intelligent Technology Using Edge Nodes .....</b>	349
Abhinav Yadav, Marushika Shukla, Tiansheng Yang, Rajkumar Singh Rathore, and Hrudaya Kumar Tripathy	
<b>Tuberculosis Disease Detection: Comparative Analysis of Logistic Regression and Decision Tree Models for Predicting TB Positivity Using Demographic and Symptom Data .....</b>	359
Ajay Kumar Tiwari and Alok Katiyar	
<b>Adam Wild Horse Optimization with QRNN for Academic Performance Prediction in a Blended Learning Model .....</b>	375
Omkar Agrahari, Vandana Dixit Kaushik, and Vinay Kumar Pathak	

<b>A Novel Attention Method to Process Long Trajectories' Sequences Efficiently .....</b>	395
Mohammed Abdalla, Hoda M. O. Mokhtar, Abdeltawab Hendawi, Tiansheng Yang, and Rajkumar Singh Rathore	
<b>Hepatitis C Prediction Applying Different ML Classification Algorithms .....</b>	415
Md. Boktiar Hossain, Khandoker Hoque, Mohammad Atikur Rahman, Priya Podder, and Deepak Gupta	
<b>Alzheimer's Disease Detection Using Hybrid Radial Basis Function Neural Network Integrated with Harris Hawk Optimization .....</b>	431
Nair Bini Balakrishnan, Anitha S. Pillai, and Jisha Jose Panackal	
<b>Assessing Wi-Fi Fingerprinting for Improved Indoor Positioning in Campus Settings: A Swedish University Example .....</b>	449
Rasmus Andersson, William Tagesson, and Rashid Ali	
<b>Optimization Algorithm of Blockchain Smart Contracts for Digital Economy .....</b>	457
Zhen Zang	
<b>Tomato Disease Detection: Leveraging YOLOv8.2.0 for Accurate and Efficient Solutions .....</b>	467
Hayder Mohammedqasim, Roa'a Mohammedqasem, Bilal A. Ozturk, Omar Akl, and Abdelkarim Boulahya	
<b>Data-Driven Facial Image Synthesis from Text Descriptions with Deep Fusion GANs .....</b>	481
Naveen Ananda Kumar Joseph Annaiah and Mohan Mahanty	
<b>Advanced Machine Learning Approach with Dynamic Analysis to Detect Malware in Cybersecurity Domain .....</b>	495
Shubhang Gupta, Shamim Khan, Tiansheng Yang, Rajkumar Singh Rathore, Aniket Das, and Nilamadhab Mishra	
<b>Enhanced Deepfake Detection Using Deep Learning on Large-Scale Video Data: A Fused ResNet50 and LSTM Approach .....</b>	503
Naveen Ananda Kumar Joseph Annaiah and B. Omkar Lakshmi Jagan	
<b>Advancements in Digital Pathology: A Comprehensive Survey of Predictive Models for Cancer Diagnosis .....</b>	517
K. Amuthachenthiru and M. Kaliappan	
<b>Detection of Coal Miner with a Comprehensive Dataset Using Transfer Learning Techniques .....</b>	535
Md. Sazedur Rahman, Khandoker Hoque, Md. Boktiar Hossain, Denesh Das, and Tao Wu	

Contents	xvii
<b>A Method for Generating a Computer Simulation Model for Detecting Scoliosis Images .....</b>	549
Yijun Zhang, Huanxiang Ding, and Jifeng Zhou	
<b>Academic Performance in Blended Learning Environment Utilizing MOOCs .....</b>	561
Anupriya Sharma Ghai, Ramesh Chander Sharma, Sanjay Jasola, and Alin Zamfiroiu	
<b>Efficient Model Used for IoT-Based Digital Forensics Using Blockchain .....</b>	577
Esha Tripathi, Upendra Kumar, Surya Prakash Tripathi, and Abhay Kumar Tripathi	
<b>Addiction-Based Tenuous Community Detection in Online Social Media .....</b>	593
Zain Ul Abideen, Fakhar Shahzad, Bharati Rathore, and Tiansheng Yang	
<b>Global Path Planning Based on Improved Ant Colony Optimization Algorithm .....</b>	615
Ruoyu Li, Jiangwen Deng, and Kaijin Qiu	
<b>Advanced Automated System for Optimal Management in Public Health Emergencies .....</b>	629
Shakil Muhammad, Adnan Mujahid Khan, Azka Qureshi, Rajakumar Arul, and Kalaipriyan Thirugnanasambandam	
<b>Machine Learning Model for Cervical Cancer Risk Assessment .....</b>	639
Vedavati Patil, Virendra Kumar Shrivastava, and Ashvini Alashetty	
<b>H<sub>∞</sub>-Based Fractional-Order Controller for Trajectory Tracking Control of Nonholonomic Mobile Robots .....</b>	651
Km Shelly Chaudhary and Naveen Kumar	
<b>MediSentBot—Medicine Review Sentiment Analysis and Recommendation Bot Using Modern NLP .....</b>	663
Shreya Rajpal and E. S. Madhan	
<b>Cybercrime Classification and Tracking Computations System Using Machine Learning .....</b>	679
Ch. Rupa, Sree Vardhan Sunkara, Rakesh Kota, G. Thippa Reddy, and Pratik Vyas	
<b>Prognostic Analysis of Logistics Operation by Modeling Ship Berthing Problem .....</b>	693
Phuoc Quy Phong Nguyen, Hoang Phuong Nguyen, Van Phuc Nguyen, Duc Chuan Nguyen, and Dang Khoa Pham Nguyen	

<b>A Virtual Cognitive Intelligence Framework for Digital Telehealth Zone</b> .....	707
Rohit Agarwal, Amit Kumar Sinha, Utpala Dutta, Lu Wang, and Bharati Rathore	
<b>Quality Detection Model of Nutmeg (<i>Myristica Fragrans Houtt</i>) Using You Only Look Once (YOLO)</b> .....	719
Manuel Soares Dos Reis Pacheco, Hadiyanto Hadiyanto, and Ridwan Sanjaya	
<b>Development of New Monolithic Zeolite Chitosan Hydrogel with Preliminary Adsorption Studies to Remove Organic Dyes from Aqueous Solutions</b> .....	731
Ghassan Abbas Alwan, Hamida Idan Salman, and Asim A. Balakit	

# Editors and Contributors

## About the Editors

**Dr. Akshi Kumar** is Senior Lecturer (Associate Professor) and Director-post graduate Research (PGR) in the Department of Computing at Goldsmiths, University of London, UK. Her academic path spans from Doctorate in the domain of web mining at the University of Delhi, India, to Post-Doctoral on explainable AI for NLP from Brazil. She held pivotal positions at Manchester Metropolitan University, UK, as well as at NSUT and DTU in New Delhi, India. Recognized as an “Exceptional Promise in AI and Data Science” by the Royal Academy of Engineering, UK, in 2022, she has received international accolades, including being named in Stanford’s “Top 2% Scientist” list in 2021 and 2022, respectively. She has successfully guided 6 doctorates and 33 master theses. Serving as Associate and Guest Editor for impactful journals, her research encompasses affective computing, social media and network analytics, natural language processing (NLP), and AI-driven health care.

**Prof. (Dr.) Abhishek Swaroop** completed his B.Tech. (CSE) from GBP University of Agriculture and Technology, M.Tech. from Punjabi University Patiala, and Ph.D. from NIT Kurukshetra. He has 28 years of teaching and industrial experience. He has served in reputed educational institutions such as Jaypee Institute of Information Technology, Noida, Sharda University Greater Noida, and Galgotias University Greater Noida. He is actively engaged in research. One of his Ph.D. scholars has completed his Ph.D. from NIT Kurukshetra, and he is currently supervising 4 Ph.D. students. He has guided 12 M.Tech. Dissertations also. He has authored 3 books and 5 book chapters. His 10 papers are indexed in DBLP and 16 papers are SCI. He had been part of the organizing committee of three IEEE conferences (ICCCA-2015, ICCCA-2016, and ICCCA-2017) and one Springer conference (ICICC-2018) as Technical Program Chair. He is Member of various professional societies like CSI and ACM and editorial board of various reputed journals.

**Dr. Pancham Shukla** is Principal Teaching Fellow in the Faculty of Engineering at Imperial College London. Dr. Shukla holds more than 25 years of teaching and research experience at various Universities in the UK and India. He holds Ph.D. in electrical and electronic engineering from Imperial College London. Dr. Pancham Shukla has supervised a broad range of projects and published papers in signal/image processing, electronic and communications systems. He has organized conferences and has served as Technical Session Chair along with reviewing many international conference and journal papers. Dr. Shukla received Vice Chancellor's award for outstanding contribution to learning and teaching as well as Student Union's award as Outstanding Member of academic staff while working at London Metropolitan University.

## Contributors

**Ali J. Abboud** Department of Computer Engineering, University of Diyala, Diyala, Baqubah, Iraq

**Mohammed Abdalla** Faculty of Computers and Artificial Intelligence, Beni-Suef University, Beni Suef, Egypt

**Azmi Shawkat Abdulbaqi** Renewable Energy Research Center, University of Anbar, Ramadi, Anbar, Iraq

**Zain Ul Abideen** Automotive Engineering Research Institute, Jiangsu University, Zhenjiang, Jiangsu, China

**Rohit Agarwal** Kalinga Institute of Industrial Technology, Deemed to Be University, Bhubaneswar, India

**Omkar Agrahari** Department of Computer Applications, School of Engineering and Technology (Formerly known as UIET Kanpur), Chhatrapati Shahu Ji Maharaj University, Uttar Pradesh, Kalyanpur, Kanpur, Uttar Pradesh, India

**Marco Aiello** Service Computing, IAAS, University of Stuttgart, Stuttgart, Germany

**Omar Akl** Faculty of Engineering, Istanbul Aydin University, Istanbul, Turkey

**Zainab H. Al-Araji** Department of Computer Science, College of Computer Science and Mathematics, University of Kufa, Najaf, Iraq

**Abdullah S. A. L.-Malaise AL-Ghamdi** Information Systems Department, School of Engineering, Computing and Design, Dar Al-Hekma University, Jeddah, Saudi Arabia

**Ashvini Alashetty** COE in Computer Vision, Department of Computer Science and Engineering, Alliance School of Advanced Computing, Alliance University, Bangalore, India

**Warqaa Shaher Alazawee** Department of Computer Engineering, University of Diyala, Diyala, Baqubah, Iraq

**Rashid Ali** Department of Engineering Science, University West, Trollhättan, Sweden

**Ghassan Abbas Alwan** Department of Chemistry, University of Kerbala, Kerbala, Iraq

**Hasanen Alyasiri** Department of Computer Science, College of Computer Science and Mathematics, University of Kufa, Najaf, Iraq

**K. Amuthachenthiru** Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamil Nadu, India

**Rasmus Andersson** Department of Engineering Science, University West, Trollhättan, Sweden

**Naveen Ananda Kumar Joseph Annaiah** Tekinvaderz LLC, Florida, USA

**Rajakumar Arul** Centre for Smart Grid Technologies, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai Campus, Chennai, India

**Ali bin Asghar** Faculty of Engineering, Istanbul Aydin University, Istanbul, Turkey

**H. B. Asif Mohamed** Department of CSE, Presidency University, Itgalpur, Rajanakunte, Yelahanka, Bengaluru, Karnataka, India

**Mona Hassan Asiri** Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

**Asim A. Balakit** College of Pharmacy, University of Babylon, Babylon, Iraq

**Nair Bini Balakrishnan** Department of Computer Applications, Hindustan Institute of Technology and Science, Chennai, India

**Satya Sai Bharath Vemula** Purdue University, West Lafayette, IN, USA

**Harsh Bhasin** Manav Rachna International Institute of Research and Studies, Faridabad, Haryana, India;

School of Computer Science Engineering & Technology, Bennett University, Greater Noida, India

**Nara Bhavyasree** Department of Computer Science Engineering, Amrita School of Computing, Bengaluru Amrita Vishwa Vidyapeetham, Bengaluru, India

**P. V. Bhuvaneshwari** Department of CSE, Presidency University, Itgalpur, Rajanakunte, Yelahanka, Bengaluru, Karnataka, India

**Abdelkarim Boulahya** Faculty of Engineering, Istanbul Aydin University, Istanbul, Turkey

**J. Mary Catherine** Computer Science, Vels Institute of Science, Technology and Advanced Studies, Chennai, Tamilnadu, India

**Akash Chandrakar** Kalinga Institute of Industrial Technology, Deemed to be University, Bhubaneswar, India

**Chandsi** Manav Rachna International Institute of Research and Studies, Faridabad, Haryana, India

**Km Shelly Chaudhary** Meerut College, Meerut, Uttar Pradesh, India; National Institute of Technology, Kurukshetra, Haryana, India

**Amit Chhabra** Department of Computer Science and Engineering, Chandigarh College of Engineering and Technology, Chandigarh, India

**Mohsin Furkh Dar** University of Hyderabad, Hyderabad, India

**Aniket Das** Kalinga Institute of Industrial Technology, Bhubaneswar, India

**Denesh Das** Department of Electrical and Computer Engineering, Lamar University, Beaumont, TX, USA

**Jiangwen Deng** Institute of Deep Learning, Southwest University, Chongqing, China

**V. Dinesh Reddy** Service Computing, IAAS, University of Stuttgart, Stuttgart, Germany;  
CSE, SRM University, Amaravati, AP, Germany

**Huanxiang Ding** School of Physical Education and Health, Linyi University, Linyi, Shandong Province, China

**Manuel Soares Dos Reis Pacheco** Diponegoro University, Semarang, Indonesia

**Utpala Dutta** Kalinga Institute of Industrial Technology, Deemed to Be University, Bhubaneswar, India

**Layth Mhmod Farhan** Faculty of Engineering, Istanbul Aydin University, Istanbul, Turkey

**Ayman G. Fayoumi** Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

**Tingsen Gan** WeBank Institute of Fintech, Shenzhen University, Shenzhen, Guangdong, China

**Avatharam Ganivada** University of Hyderabad, Hyderabad, India

**Anupriya Sharma Ghai** School of Computing, Graphic Era Hill University, Dehradun, India

**Deepak Gupta** Department of Computer Science and Engineering, Maharaja Agrasen Institute of Technology, Delhi, India

**Hardik Gupta** Department of Computer Science and Engineering, Chandigarh College of Engineering and Technology, Chandigarh, India

**Shubhang Gupta** Kalinga Institute of Industrial Technology, Bhubaneswar, India

**Hadiyanto Hadiyanto** Diponegoro University, Semarang, Indonesia

**Habib Rahman Hamedy** Faculty of Engineering, Istanbul Aydin University, Istanbul, Turkey

**Abdeltawab Hendawi** Department of Computer Science and Statistics, University of Rhode Island, Kingston, USA

**Khandoker Hoque** School of Engineering, San Francisco Bay University, Fremont, CA, USA

**Md. Boktiar Hossain** School of Engineering, San Francisco Bay University, Fremont, CA, USA

**Abualqasim Khalil Ismael** Faculty of Engineering, Istanbul Aydin University, Istanbul, Turkey

**B. Omkar Lakshmi Jagan** Department of Electrical and Electronics Engineering, Vignan's Institute of Information Technology, Duvvada, Visakhapatnam, Andhra Pradesh, India;  
Department of Computer Science and Engineering, Vignan's Institute of Information Technology, Duvvada, Visakhapatnam, Andhra Pradesh, India

**Sanjay Jasola** Department of Computer Science and Engineering, Graphic Era Hill University, Dehradun, India

**M. Kaliappan** Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamil Nadu, India

**Ananth Mahesh Kashyap** University of Stuttgart, Stuttgart, Germany

**Alok Katiyar** School of Computer Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India

**Vandana Dixit Kaushik** Professor, Department of Computer Science and Engineering, Harcourt Butler Technical University, Uttar Pradesh, Nawabganj, Kanpur, Uttar Pradesh, India

**Shamim Khan** Kalinga Institute of Industrial Technology, Bhubaneswar, India

**Charan Ramtej Kodi** University of Hyderabad, Hyderabad, Telangana, India

**Rakesh Kota** Velagapudi Ramakrishna Siddhartha Engineering College, Kanuru, Vijayawada, India

**R. Harihara Krishnan** Computer Application, Patrician College of Arts and Science, Chennai, India

**Varsha Venkata Krishnan** Purdue University, West Lafayette, IN, USA

**R. Krishnaprasanna** Department of Electronics and Communication Engineering, Sathyabama Institute of Science and Technology, Chennai, India

**Ankit Kumar** Kalinga Institute of Industrial Technology, Bhubaneswar, India

**Bijendra Kumar** Department of Computer Science and Engineering, NSUT, Delhi, India

**Kapila Kumar** Manav Rachna International Institute of Research and Studies, Faridabad, Haryana, India

**Lavish Kumar** Department of Software Engineering, Delhi Technological University (DTU), Delhi, New Delhi, India

**Naveen Kumar** Mahatma Jyotiba Phule Rohilkhand University Bareilly, Bareilly, Uttar Pradesh, India;

National Institute of Technology, Kurukshetra, Haryana, India

**Sudhakar Kumar** Department of Computer Science and Engineering, Chandigarh College of Engineering and Technology, Chandigarh, India

**Upendra Kumar** Institute of Engineering and Technology, Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India

**Ritika Kumari** Department of Artificial Intelligence and Data Sciences, IGDTUW, New Delhi, Delhi, India

**Ruoyu Li** School of Communication, Hong Kong Baptist University, Hong Kong, China

**Changgui Lin** Gansu Agricultural University, Anning District, Lanzhou, Gansu Province, China

**Anjia Ma** College of Design and Art, Shandong Huayu University of Technology, Dezhou, Shandong, China

**E. S. Madhan** Vellore Institute of Technology, Vellore, Tamil Nadu, India

**Mohan Mahanty** Department of Computer Science and Engineering, Vignan's Institute of Information Technology, Duvvada, Visakhapatnam, Andhra Pradesh, India

**Raghda Salam Al. Mahdawi** Department of Computer Engineering, University of Diyala, Diyala, Baqubah, Iraq

**Manya** Department of Computer Science and Engineering, Chandigarh College of Engineering and Technology, Chandigarh, India

**Shweta Meena** Department of Software Engineering, Delhi Technological University (DTU), Delhi, New Delhi, India

**Yan Meng** School of Foreign Languages, Huainan Normal University, Huainan, China

**Nilamadhab Mishra** VIT Bhopal University, Sehore, Madhya Pradesh, India

**Roa'a Mohammedqasem** Faculty of Engineering, Istanbul Aydin University, Istanbul, Turkey

**Hayder Mohammedqasim** Faculty of Engineering, Istanbul Aydin University, Istanbul, Turkey

**Rishabh Mohata** Kalinga Institute of Industrial Technology, Deemed to be University, Bhubaneswar, India

**Hoda M. O. Mokhtar** Egypt University of Informatics, Cairo, Egypt

**Shakil Muhammad** Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea

**Adnan Mujahid Khan** History Ltd, Coventry, UK

**Bharath Goud Musalaya** University of Central Missouri, Warrensburg, MO, USA

**Dang Khoa Pham Nguyen** Institute of Maritime, Ho Chi Minh City University of Transport, Ho Chi Minh City, Vietnam

**Duc Chuan Nguyen** Institute of Maritime, Ho Chi Minh City University of Transport, Ho Chi Minh City, Vietnam

**Hoang Phuong Nguyen** Academy of Politics Region II, Ho Chi Minh City, Vietnam

**Phuoc Quy Phong Nguyen** Faculty of Maritime, Maritime College II, Ho Chi Minh City, Vietnam

**Van Phuc Nguyen** Institute of Maritime, Ho Chi Minh City University of Transport, Ho Chi Minh City, Vietnam

**Nagabhryu Nikitha** Department of Computer Science Engineering, Amrita School of Computing, Bengaluru Amrita Vishwa Vidyapeetham, Bengaluru, India

**Bilal A. Ozturk** Faculty of Engineering, Istanbul Aydin University, Istanbul, Turkey

**Jisha Jose Panackal** Department of Computer Science, Sacred Heart College, Thrissur, Kerala, India

**Vinay Kumar Pathak** Professor, Department of Computer Science and Engineering, Harcourt Butler Technical University, Uttar Pradesh, Nawabganj, Kanpur, Uttar Pradesh, India;

Professor, Chhatrapati Shahu Ji Maharaj University, Kalyanpur, Kanpur, Uttar Pradesh, India

**Vedavati Patil** COE in Computer Vision, Department of Computer Science and Engineering, Alliance School of Advanced Computing, Alliance University, Bangalore, India

**Anitha S. Pillai** Department of Computer Applications, Hindustan Institute of Technology and Science, Chennai, India

**Prajoy Podder** Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO, USA

**Priya Podder** Dhaka National Medical College, Dhaka, Bangladesh

**V. Poornima** Chevalier T. Thomas Elizabeth College for Women, Chennai, Tamilnadu, India

**S. Pratheepa** Department of Computer Science, J.H.A. Agarsen College, Madhavaram, Chennai, Tamilnadu, India

**Nikhil Kumar Pulipeta** University of Central Missouri, Warrensburg, MO, USA

**Kaijin Qiu** College of Computer and Information Science, Southwest University, Chongqing, China

**Azka Qureshi** Applab Qatar, Doha, Qatar

**D. Radha** Department of Computer Science Engineering, Amrita School of Computing, Bengaluru Amrita Vishwa Vidyapeetham, Bengaluru, India

**Mahmoud Ragab** Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

**Abdul Rahim** Department of Computer Science and Engineering, Chandigarh College of Engineering and Technology, Chandigarh, India

**Md. Sazedur Rahman** Department of Computer Science, Missouri University of Science and Technology, Rolla, MO, USA

**Mohammad Atikur Rahman** School of Engineering, San Francisco Bay University, Fremont, CA, USA;  
Institute of Information and Communication Technology, BUET, Dhaka, Bangladesh

**Aaryan Raj** Kalinga Institute of Industrial Technology, Deemed to be University, Bhubaneswar, India

**Vishal Rajkumar** Purdue University, West Lafayette, IN, USA

**Shreya Rajpal** Vellore Institute of Technology, Vellore, Tamil Nadu, India

**Bharati Rathore** University of South Wales, Pontypridd, UK

**Rajkumar Singh Rathore** Cardiff School of Technologies, Cardiff Metropolitan University, Cardiff, UK

**T. Ravi** Department of Electronics and Communication Engineering, Sathyabama Institute of Science and Technology, Chennai, India

**G. Thippa Reddy** The College of Mathematics and Computer Science, Zhejiang University, Hangzhou, China;

Division of Research and Development, Lovely Professional University, Phagwara, India;

Center of Research Impact and Outcome, Chitkara University, Rajpura, Punjab, India

**Ch. Rupa** Velagapudi Ramakrishna Siddhartha Engineering College, Kanuru, Vijayawada, India

**Hamida Idan Salman** Department of Chemistry, University of Kerbala, Kerbala, Iraq

**Ridwan Sanjaya** Soegijapranata Catholic University, Semarang, Indonesia

**Khushi Saxena** Kalinga Institute of Industrial Technology, Bhubaneswar, India

**Fakhar Shahzad** Research Institute of Business Analytics and Supply Chain Management, College of Management, Shenzhen University, Shenzhen, China

**Karan Sharma** Department of Computer Science and Engineering, Chandigarh College of Engineering and Technology, Chandigarh, India

**Ramesh Chander Sharma** Department of Computer Science and Engineering, Graphic Era Hill University, Dehradun, India

**Shilpi Sharma** Department of Computer Science and Engineering, NSUT, Delhi, India

**Virendra Kumar Shrivastava** COE in Computer Vision, Department of Computer Science and Engineering, Alliance School of Advanced Computing, Alliance University, Bangalore, India

**Marushika Shukla** Kalinga Institute of Industrial Technology, Deemed to Be University, Bhubaneswar, India

**Sunil K. Singh** Department of Computer Science and Engineering, Chandigarh College of Engineering and Technology, Chandigarh, India

**Amit Kumar Sinha** Kalinga Institute of Industrial Technology, Deemed to Be University, Bhubaneswar, India

**Ms. Soumya** Department of Computer Science Engineering, School of Computer Science Engineering and Information Science, Presidency University, Bengaluru, Karnataka, India

**Saksham Srivastava** Kalinga Institute of Industrial Technology, Bhubaneswar, India

**Ruikai Sun** Cardiff University, Cardiff, UK

**Sree Vardhan Sunkara** Velagapudi Ramakrishna Siddhartha Engineering College, Kanuru, Vijayawada, India

**Gunapati Suresh** Sathyabama Institute of Science and Technology, Chennai, India

**William Tagesson** Department of Engineering Science, University West, Trollhättan, Sweden

**V. Thamilarasi** Department of Computer Science, Sri Sarada College for Women (Autonomous), Salem, Tamilnadu, India

**Kalaipriyan Thirugnanasambandam** Centre for Smart Grid Technologies, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai Campus, Chennai, India

**Ajay Kumar Tiwari** School of Computer Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India

**Abhay Kumar Tripathi** Pranveer Singh Institute of Technology, Kanpur, India

**Esha Tripathi** Institute of Engineering and Technology, Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India;  
Pranveer Singh Institute of Technology, Kanpur, India

**Surya Prakash Tripathi** R R Institute of Modern Technology, Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India

**Hrudaya Kumar Tripathy** Kalinga Institute of Industrial Technology, Deemed to be University, Bhubaneswar, India

**Pankhuri Verma** Service Computing, IAAS, University of Stuttgart, Stuttgart, Germany

**Radhakrishnan Vignesh** Department of CSE, Presidency University, Itgalpur, Rajanakunte, Yelahanka, Bengaluru, Karnataka, India

**V. Vijayalakshmi** Department of Computer Science, Christ College of Science and Management, Sonnur, Karnataka, India

**Kogatam Vijayasree** Department of Computer Science Engineering, Amrita School of Computing, Bengaluru Amrita Vishwa Vidyapeetham, Bengaluru, India

**Robin Rohit Vincent** Computer Science Engineering, Presidency University, Bengaluru, Karnataka, India

**Pratik Vyas** Department of Computer Science, Nottingham Trent University, Nottingham, UK

**Lu Wang** Xi'an Jiaotong-Liverpool University, Wuzhong District, Suzhou, China

**Wei Wang** WeBank Institute of Fintech, Shenzhen University, Shenzhen, Guangdong, China;

Shenzhen Audencia Financial Technology Institute, Shenzhen University,  
Shenzhen, Guangdong, China

**Tao Wu** Department of Computer Science, Missouri University of Science and  
Technology, Rolla, MO, USA

**Abhinav Yadav** Kalinga Institute of Industrial Technology, Deemed to Be  
University, Bhubaneswar, India

**Kelang Yang** WeBank Institute of Fintech, Shenzhen University, Shenzhen,  
Guangdong, China

**Tiansheng Yang** University of South Wales, Pontypridd, UK

**Noor Yeshfeen** Department of Artificial Intelligence and Data Sciences,  
IGDTUW, New Delhi, Delhi, India

**Alin Zamfirou** Bucharest University of Economic Studies, Bucharest, Romania;  
National Institute for Research and Development in Informatics—ICI Bucharest,  
Bucharest, Romania

**Zhen Zang** School of Economics and Management, Chengdu Technological  
University, Chengdu, China

**Maciej Zawodniok** Department of Electrical and Computer Engineering,  
Missouri University of Science and Technology, Rolla, MO, USA

**Jiuquan Zhang** School of Foreign Languages, Huainan Normal University,  
Huainan, China

**Yijun Zhang** School of Physical Education and Health, Linyi University, Linyi,  
Shandong Province, China

**Jifeng Zhou** School of Physical Education and Health, Linyi University, Linyi,  
Shandong Province, China

**Soukeina Zouaidi** ESPRIT School of Engineering, Tunis, Tunisia

# Dynamic Weight-Adjusted Ensemble Loss for Enhanced Medical Image Segmentation



Mohsin Furkh Dar<sup>✉</sup> and Avatharam Ganivada<sup>✉</sup>

**Abstract** The choice of loss function is crucial in medical image segmentation, as it directly influences the accuracy and robustness of the model. Conventional loss functions, such as Binary Cross-Entropy (BCE) and Dice loss, often suffer from imbalanced optimization, focusing too much on either region overlap or boundary accuracy, leading to suboptimal segmentation results. To address this limitation, we propose a dynamic weight-adjusted ensemble loss, called as DyWAEn loss, function that combines BCE, Dice, Hausdorff, and Tversky losses. Our method dynamically adjusts the weights of these loss functions based on their performance during training, allowing the model to emphasize the most relevant losses. This approach effectively mitigates the shortcomings of traditional loss functions by providing a balanced optimization across various metrics. Experimental results on the ultrasound and polyp image datasets demonstrate that the DyWAEn loss significantly outperforms individual loss functions in terms of Dice coefficient and Intersection-Over-Union (IoU). The DyWAEn loss function is highly adaptable and can be seamlessly incorporated into a wide range of segmentation frameworks without the need for adjustments, providing a flexible and effective solution for improving deep learning models in medical imaging.

**Keywords** Segmentation · Deep learning · Medical imaging · Loss functions

---

M. F. Dar (✉) · A. Ganivada  
University of Hyderabad, Hyderabad, India  
e-mail: [20mcpc02@uohyd.ac.in](mailto:20mcpc02@uohyd.ac.in)

A. Ganivada  
e-mail: [avatharg@uohyd.ac.in](mailto:avatharg@uohyd.ac.in)

## 1 Introduction

Medical image segmentation is a critical task in medical imaging that significantly impacts diagnosis, treatment planning, and disease monitoring [7]. Deep learning models, particularly convolutional neural networks (CNNs) such as UNet [14], have become the standard approach for this task due to their powerful feature extraction and localization capabilities. However, the choice of loss function remains a crucial factor influencing the model's performance, especially in challenging tasks like medical image segmentation where precision is paramount [2].

Dice loss and Binary Cross-Entropy (BCE) loss are popular choices for medical image segmentation tasks, offering simplicity and reasonable performance. Nevertheless, these loss functions have limitations, including their inability to effectively handle class imbalance and boundary errors [5]. BCE loss, for example, focuses on pixel-wise classification and often fails to capture spatial relationships within the image, leading to suboptimal segmentation, especially for small or elongated structures [4]. Dice loss, although useful for dealing with class imbalances by emphasizing overlap between predicted and ground truth masks, can be sensitive to boundary errors [11]. To address these issues, more sophisticated loss functions have been proposed. Hausdorff loss prioritizes boundary accuracy by calculating the maximum distance between the predicted and ground truth mask boundaries [10]. Tversky loss generalizes Dice loss to handle imbalanced datasets by allowing adjustable parameters to balance false positives and false negatives [15]. Despite their individual strengths, these loss functions still face challenges when applied independently.

Researchers have sought to leverage the strengths of various loss functions while mitigating their individual weaknesses by combining them [9, 13, 16, 17]. While these methods demonstrated improvements, they also introduced additional hyperparameters to balance the contributions of each loss function. This added complexity often necessitated extensive tuning, complicating the optimization process [3, 6]. To address these limitations, we propose a DyWAEn loss function that combines BCE, Dice, Hausdorff, and Tversky losses.

The proposed DyWAEn loss dynamically adjusts the weights of individual loss functions based on their performance during training, ensuring balanced optimization across various metrics. This approach mitigates the need for extensive hyperparameter tuning and provides a more transparent evaluation of each loss function's contribution. By introducing a dynamic weighting mechanism, our approach effectively balances the contributions of multiple loss functions, thereby overcoming the individual limitations of conventional loss functions. The DyWAEn loss function is highly adaptable, allowing it to be seamlessly integrated into a wide range of segmentation frameworks without requiring any changes to the network architecture or data preprocessing steps. This makes it a valuable tool for enhancing deep learning tasks in medical imaging.

The rest of the paper is organized as follows: Sect. 2 reviews related work on loss functions for medical image segmentation. Section 3 details the proposed DyWAEn loss and its components. Section 4 describes the experimental setup, including the

dataset, training procedure, and implementation details. Section 5 presents the experimental results and discusses the performance of the proposed method compared to existing approaches. Finally, Sect. 6 concludes the paper.

## 2 Related Work

In medical image segmentation, the selection of an appropriate loss function is pivotal to achieving high-performance models. Various loss functions have been explored in the literature, each addressing specific challenges such as class imbalance, boundary precision, and the need for robust generalization across diverse datasets. While traditional loss functions like BCE and Dice loss have been widely used, their limitations in handling class imbalance and boundary errors have motivated researchers to explore more sophisticated alternatives.

To address these challenges, various loss functions have been proposed. Hausdorff loss, which focuses on improving boundary accuracy, has been explored in [10]. Tversky loss, an extension of Dice loss that handles imbalanced datasets, was introduced in [15].

Combining multiple loss functions to leverage their individual strengths has been a growing trend in the field. Taghanaki et al. [16] introduced a combination of Dice and cross-entropy losses to achieve a balance between class overlap and pixel-wise accuracy. Wu et al. [17] further extended this idea by integrating cross-entropy loss with spatial edge maps, resulting in improved segmentation of complex anatomical structures. Similarly, Peng et al. [13] proposed a composite loss function that merges Dice and focal losses to enhance the robustness of segmentation models across various tasks. While these approaches have demonstrated improvements in performance, they often introduce additional hyperparameters, which can complicate the training process and require extensive tuning to achieve optimal results.

Recent advancements have focused on dynamic weighting strategies that automatically adjust the contributions of individual loss functions during training. These methods aim to reduce the need for manual hyperparameter tuning while ensuring balanced optimization across different metrics. For example, Kamran et al. [9] proposed a weighted ensemble loss that adjusts the weights of Hinge loss and Mean Squared Error based on their performance throughout the training process. Despite the promising results, these methods still require some level of manual intervention to fine-tune the weighting strategy, which can be challenging in practice.

In light of these developments, the proposed DyWAE<sub>n</sub> loss function builds upon the existing body of work by introducing a dynamic weighting mechanism that automatically balances the contributions of BCE, Dice, Hausdorff, and Tversky losses. This approach not only simplifies the optimization process but also enhances segmentation performance across a variety of challenging medical imaging tasks. The DyWAE<sub>n</sub> loss function represents a significant step forward in the design of robust and adaptable loss functions for medical image segmentation.

### 3 Methodology

In this study, we propose a DyWAEn loss function that combines multiple conventional loss functions to enhance the performance of medical image segmentation. The DyWAEn loss function integrates Binary Cross-Entropy (BCE) loss, Dice loss, Hausdorff loss, and Tversky loss, dynamically adjusting their weights during training to ensure balanced optimization across different metrics. This section details the formulation and implementation of the DyWAEn loss function.

#### 3.1 *Binary Cross-Entropy (BCE) Loss*

Binary Cross-Entropy loss is commonly used for binary classification tasks and is defined as:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (1)$$

where  $y_i$  is the ground truth label for pixel  $i$ ,  $p_i$  is the predicted probability for pixel  $i$ , and  $N$  is the total number of pixels. BCE loss effectively handles pixel-wise classification but often fails to capture spatial relationships.

#### 3.2 *Dice Loss*

Dice loss measures the overlap between the predicted segmentation and the ground truth, which is particularly useful for handling imbalanced classes. It is defined as:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2 \sum_{i=1}^N y_i p_i}{\sum_{i=1}^N y_i + \sum_{i=1}^N p_i} \quad (2)$$

This loss function helps improve segmentation accuracy for small structures but can be sensitive to boundary errors.

#### 3.3 *Hausdorff Loss*

Hausdorff loss measures the maximum distance between the boundary points of the predicted and ground truth masks, improving boundary accuracy. It is formulated as:

$$\mathcal{L}_{\text{Hausdorff}} = \max \left( \sup_{x \in \partial X} \inf_{y \in \partial Y} d(x, y), \sup_{y \in \partial Y} \inf_{x \in \partial X} d(x, y) \right) \quad (3)$$

where  $\partial X$  and  $\partial Y$  are the boundaries of the predicted and ground truth masks, respectively, and  $d(x, y)$  is the Euclidean distance between points  $x$  and  $y$ .

### 3.4 Tversky Loss

Tversky loss extends Dice loss by introducing parameters  $\alpha$  and  $\beta$  to control the balance between false positives and false negatives. It is defined as:

$$\mathcal{L}_{\text{Tversky}} = 1 - \frac{\sum_{i=1}^N y_i p_i}{\sum_{i=1}^N y_i p_i + \alpha \sum_{i=1}^N y_i (1 - p_i) + \beta \sum_{i=1}^N (1 - y_i) p_i} \quad (4)$$

This loss function offers flexibility in handling imbalanced datasets by adjusting the parameters  $\alpha$  and  $\beta$ .

### 3.5 DyWAEn Loss

The proposed DyWAEn loss function combines the aforementioned loss functions and adjusts their weights dynamically based on their performance during training. The DyWAEn loss function is defined as:

$$\begin{aligned} \mathcal{L}_{\text{DyWAEn}} = & w_{\text{BCE}} \cdot \mathcal{L}_{\text{BCE}} + w_{\text{Dice}} \cdot \mathcal{L}_{\text{Dice}} \\ & + w_{\text{Hausdorff}} \cdot \mathcal{L}_{\text{Hausdorff}} + w_{\text{Tversky}} \cdot \mathcal{L}_{\text{Tversky}} \end{aligned} \quad (5)$$

where  $w_{\text{BCE}}$ ,  $w_{\text{Dice}}$ ,  $w_{\text{Hausdorff}}$ , and  $w_{\text{Tversky}}$  are the weights assigned to each loss function. These weights are updated dynamically at the end of each epoch to reflect the performance of each loss function. The adjustment mechanism ensures that the DyWAEn loss function focuses on the most relevant loss components, leading to balanced optimization across different metrics.

The weight adjustment process involves calculating the validation loss for each component and normalizing these losses to determine the new weights. Specifically, the weights are updated as follows:

$$w_i = \frac{1 - \frac{\mathcal{L}_i}{\sum_{j=1}^M \mathcal{L}_j}}{M - 1} \quad (6)$$

where  $\mathcal{L}_i$  is the validation loss for the  $i$ -th loss function, and  $\sum_{j=1}^M \mathcal{L}_j$  is the sum of all individual validation losses, with  $M$  being the total number of loss functions. This normalization ensures that the weights sum to 1, providing a balanced contribution from each loss function.

## 4 Experimental Setup

### 4.1 Implementation

The DyWAE<sub>n</sub> loss function is implemented as a custom loss function in TensorFlow/Keras. During training, the weights are updated at the end of each epoch based on the validation performance of each loss function. This dynamic adjustment allows the model to emphasize the most relevant loss functions, thereby improving segmentation accuracy and robustness.

Two deep learning models, UNet [14] and Attention UNet [12], are used to implement and evaluate the DyWAE<sub>n</sub> loss function.

**UNet Model** UNet is a widely used convolutional neural network architecture designed for biomedical image segmentation [14]. It features an encoder–decoder structure with skip connections that link corresponding layers of the encoder and decoder, enabling the model to effectively capture both spatial and contextual information, which is essential for segmentation tasks. In this study, the UNet model is configured with an input size of  $128 \times 128 \times 1$ , specifically for grayscale images. The encoder consists of five convolutional blocks, each followed by a max-pooling layer to progressively downsample the input and extract features. The bottleneck layer in the center of the network further processes the encoded features through a convolutional block. The decoder then reconstructs the segmented image using five upsampling blocks, each followed by concatenation with the corresponding encoder block and an additional convolutional block to refine the output. Finally, a sigmoid activation function is applied to the output layer to perform binary segmentation.

**Attention UNet Model** Attention UNet extends the UNet architecture by incorporating attention mechanisms [12]. Attention gates are added to the skip connections to emphasize relevant features and suppress irrelevant ones. This enhancement improves the model’s ability to focus on important regions, leading to better segmentation performance. The Attention UNet model used in this study has a similar configuration to the UNet model, with the addition of attention gates in the skip connections.

**Table 1** Training configuration for the segmentation models

Parameter	Value
Optimizer	Adam
Learning rate	$1 \times 10^{-4}$
Batch size	16
Epochs	100

## 4.2 Training and Evaluation

The DyWAEn loss function and the training process are designed to adaptively focus on the most relevant loss functions, thereby improving segmentation accuracy and robustness. This approach can be easily integrated into existing deep learning frameworks for medical image segmentation, offering a versatile solution for various segmentation tasks. The training configuration is summarized in Table 1.

## 4.3 Datasets

This study evaluates the DyWAEn loss on three publicly available medical image datasets, each with distinct clinical demographics. The BUSI dataset, collected at the Baheya Hospital for early detection and treatment of women's cancer in Cairo, Egypt, includes 437 benign and 210 malignant lesions imaged using LOGIQ E9 and LOGIQ E9 Agile ultrasound systems [1]. The UDIAT dataset comprises 163 BUS images from patients in Spain, containing 110 benign and 53 malignant samples [18]. The Kvasir-SEG dataset, sourced from colonoscopy procedures, includes 1000 polyp images with pixel-level segmentation masks annotated by medical experts, aiding in the development of polyp detection algorithms [8].

Each dataset consists of images and their corresponding segmentation masks. To ensure consistency and facilitate effective training of our deep learning models, we preprocess all images and masks uniformly. This preprocessing involves resizing each image and mask to a standard dimension of  $128 \times 128$  pixels and normalizing the pixel values to the  $[0, 1]$  range. This normalization aids in stabilizing the training process and ensuring standardized input data.

## 5 Results and Analysis

The effectiveness of the proposed DyWAEn loss function is evaluated using the BUSI, UDIAT, and Kvasir-SEG datasets. The performance of the DyWAEn loss

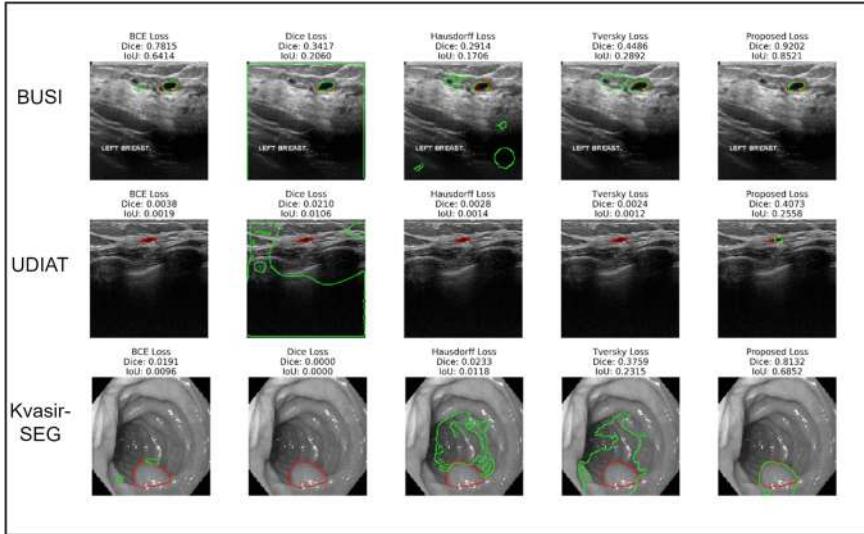
**Table 2** Comparative performance of loss functions across datasets (BUSI, UDIAT, and Kvasir-SEG) using fivefold cross-validation

Dataset	Loss function	UNet		Attention UNet	
		DSC	IoU	DSC	IoU
BUSI	BCE loss	0.7549±0.022	0.6065±0.034	0.7390±0.010	0.5868±0.053
	Dice loss	0.6668±0.018	0.5005±0.029	0.5587±0.179	0.3888±0.566
	Hausdorff loss	0.7519±0.006	0.6030±0.010	0.63551±0.460	0.4865±0.566
	Tversky loss	0.7542±0.005	0.6055±0.008	0.7482±0.040	0.5985±0.060
	DyWAEn loss	<b>0.7655±0.019</b>	<b>0.6220±0.039</b>	<b>0.7613±0.015</b>	<b>0.6179±0.034</b>
UDIAT	BCE loss	0.6598±0.115	0.4941±0.185	0.6231±0.182	0.4633±0.335
	Dice loss	0.1489±0.639	0.0811±0.679	0.2803±0.876	0.1688±0.968
	Hausdorff loss	0.6773±0.316	0.5187±0.486	0.7097±0.197	0.5552±0.344
	Tversky loss	0.7108±0.013	0.5583±0.026	0.5832±0.616	0.4302±0.818
	DyWAEn loss	<b>0.7607±0.008</b>	<b>0.6165±0.016</b>	<b>0.7500±0.051</b>	<b>0.6058±0.097</b>
Kvasir-SEG	BCE loss	0.7576±0.060	0.6093±0.093	0.7542±0.057	0.6058±0.092
	Dice loss	0.4278±0.968	0.3006±0.812	0.6812±0.165	0.5186±0.245
	Hausdorff loss	0.7656±0.005	0.6204±0.009	0.7265±0.132	0.5725±0.198
	Tversky loss	0.7668±0.008	0.6220±0.012	0.7596±0.047	0.6127±0.078
	DyWAEn loss	<b>0.7912±0.032</b>	<b>0.6547±0.054</b>	<b>0.7903±0.025</b>	<b>0.6534±0.041</b>

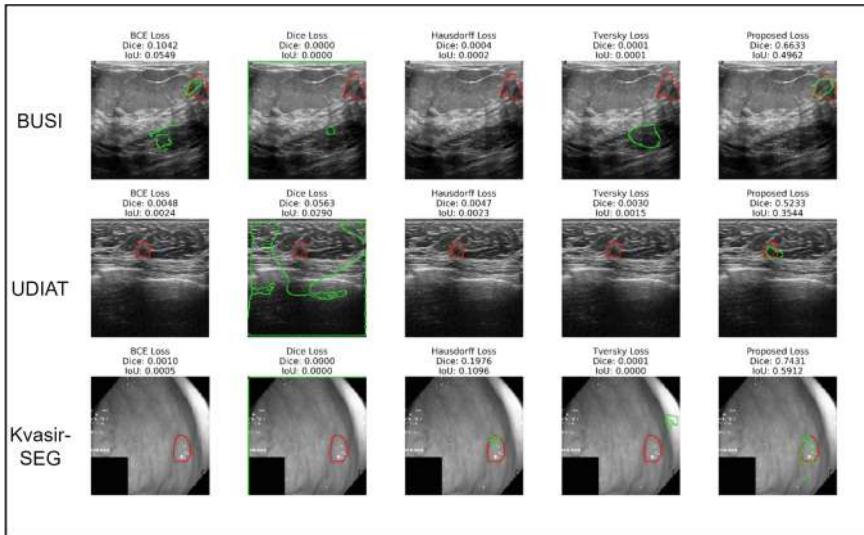
function is compared against several baseline loss functions, including Binary Cross-Entropy (BCE) loss, Dice loss, Hausdorff loss, and Tversky loss, implemented within UNet and Attention UNet models. The evaluation metrics used are Dice Similarity Coefficient (DSC) and Intersection-Over-Union (IoU), which are standard metrics for segmentation tasks.

Table 2 presents the comparative performance of the DyWAEn loss function and the baseline loss functions across the three datasets. The DyWAEn loss function consistently achieves higher DSC and IoU scores compared to the baseline loss functions. Specifically, on the Kvasir-SEG dataset, the DyWAEn loss function achieves a mean DSC of 0.7912 and a mean IoU of 0.6547 using the UNet model, and a mean DSC of 0.7903 and a mean IoU of 0.6534 using the Attention UNet model. These results highlight the superior performance of the DyWAEn loss function, particularly in challenging segmentation scenarios with ambiguous boundaries and varying lesion shapes.

Figures 1 and 2 provide a qualitative comparison of the segmentation results using UNet and Attention UNet models, respectively, with different loss functions on the BUSI, UDIAT, and Kvasir-SEG datasets. In these figures, the red contours represent the ground truth annotations, while the green contours represent the predicted segmentations. The figures illustrate that the DyWAEn loss function produces segmentations that closely match the ground truth annotations, demonstrating its ability to accurately capture the boundaries and structures of the lesions.



**Fig. 1** Qualitative comparison of segmentation results using UNet with different loss functions on three publicly available datasets (BUSI, UDIAT, and Kvasir-SEG). The red contours represent the ground truth annotations, while the green contours represent the predicted segmentations. Dice and IoU scores are shown for each method and dataset



**Fig. 2** Qualitative comparison of segmentation results using Attention UNet with different loss functions on three publicly available datasets (BUSI, UDIAT, and Kvasir-SEG). The red contours represent the ground truth annotations, while the green contours represent the predicted segmentations. Dice and IoU scores are shown for each method and dataset

**Table 3** Statistical significance ( $p$ -values) of paired  $t$ -tests comparing the Dice scores of UNet models trained with the DyWAEn loss against baseline loss functions (BCE, Dice, HD, Tversky) across three datasets (BUSI, UDIAT, and Kvasir-SEG)

Loss function	BUSI	UDIAT	Kvasir-SEG
BCE loss	0.008	0.006	0.0090
Dice loss	0.000	0.000	0.0000
HD loss	0.003	0.003	0.0009
Tversky loss	0.010	0.029	0.0094

## 5.1 Comparative Analysis

## 5.2 Statistical Significance

To assess the statistical significance of the observed improvements in segmentation performance, we performed paired  $t$ -tests comparing the Dice scores obtained with the DyWAEn loss against those obtained with each baseline loss function (BCE, Dice, Hausdorff, and Tversky) on each of the three datasets (Table 3). The results demonstrate that the DyWAEn loss consistently outperforms the baseline losses with high statistical significance ( $p < 0.05$ ) across all datasets. This finding underscores the robustness and effectiveness of the DyWAEn loss in handling the diverse challenges of medical image segmentation, including ambiguous boundaries, varying lesion shapes, and data imbalance.

The most notable differences are observed in the Kvasir-SEG dataset, where the DyWAEn loss significantly outperforms all baseline losses ( $p < 0.01$ ). This suggests that the DyWAEn loss's ability to model boundary uncertainty and handle class imbalance is particularly beneficial in these datasets, which may exhibit more challenging segmentation scenarios.

In contrast, the differences in Dice scores between the DyWAEn loss and Tversky losses are less pronounced on the BUSI and UDIAT dataset, although the DyWAEn loss still consistently achieves statistically significant improvements. This could be attributed to the flexibility of Tversky loss by adjusting  $\alpha$  and  $\beta$  to the specific requirements of the medical segmentation task at hand.

Overall, the statistical analysis confirms that the improvements achieved by the DyWAEn loss are not due to chance but represent a genuine and significant advancement in the segmentation performance. These findings further validate the DyWAEn loss as a promising tool for enhancing the accuracy and reliability of medical image segmentation, potentially leading to improved diagnosis and treatment planning of the disease.

## 6 Conclusion

This study addresses the challenge of optimizing medical image segmentation by proposing the DyWAEn loss function, which dynamically balances BCE, Dice, Hausdorff, and Tversky losses. Key findings demonstrate significant improvements in segmentation performance on BUSI, UDIAT, and Kvasir-SEG datasets, as measured by Dice coefficient and IoU. This method addresses the limitations of traditional loss functions, such as imbalanced optimization and the need for extensive hyperparameter tuning, by providing a more adaptive and comprehensive approach to loss calculation. However, this approach introduces relies on validation performance, which may limit generalizability. Future research should explore integration with other network architectures, applicability to diverse imaging modalities, and development of more efficient weight adjustment strategies.

## References

1. Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A (2020) Dataset of breast ultrasound images. *Data Brief* 28:104863
2. Ansari MA, Mangalote IAC, Meher PK, Aboumarzouk O, Al-Ansari A, Halabi O, Dakua SP (2024) Advancements in deep learning for b-mode ultrasound segmentation: a comprehensive review. *IEEE Trans Emerg Top Comput Intell* 8(3):2126–2149
3. Bondugula RK, Bommi NS, Udgata SK (2024) An efficient multi-stage ensemble deep learning framework for diagnosing infectious diseases. *Decis Anal J* 11:100458
4. Chen Y, Yu L, Wang J-Y, Panjwani N, Obeid J-P, Liu W, Liu L, Kovalchuk N, Gensheimer MF, Vitzthum LK, Beadle BM, Chang DT, Le Q-T, Han B, Xing L (2023) Adaptive region-specific loss for improved medical image segmentation. *IEEE Trans Pattern Anal Mach Intell* 45(11):13408–13421
5. Dar MF, Ganivada A (2023) Efficient U-Net: a novel deep learning method for breast tumor segmentation and classification in ultrasound images. *Neural Process Lett* 55:10439–10462
6. Dar MF, Ganivada A (2024) Deep learning and genetic algorithm-based ensemble model for feature selection and classification of breast ultrasound images. *Image Vis Comput* 146:105018
7. Du J, Guan K, Liu P, Li Y, Wang T (2023) Boundary-sensitive loss function with location constraint for hard region segmentation. *IEEE J Biomed Health Inf* 27(2):992–1003
8. Jha D, Smedsrød PH, Riegler MA, Halvorsen P, de Lange T, Johansen D, Johansen HD (2020) Kvasir-seg: a segmented polyp dataset. In: Ro YM, Cheng W-H, Kim J, Chu W-T, Cui P, Choi J-W, Hu M-C, De Neve W (eds) MultiMedia modeling. Springer International Publishing, Cham, pp 451–462
9. Kamran SA, Hossain KF, Tavakkoli A, Zuckerbrod SL, Sanders KM, Baker SA (2021) Rv-gan: segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial network. In: de Bruijne M, Cattin PC, Cotin S, Padov N, Speidel S, Zheng Y, Essert C (eds) Medical image computing and computer assisted intervention (MICCAI 2021). Springer International Publishing, Cham, pp 34–44
10. Karimi D, Salcudean SE (2020) Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Trans Med Imaging* 39(2):499–513
11. Lin Q, Chen X, Chen C, Garibaldi JM (2024) Boundary-wise loss for medical image segmentation based on fuzzy rough sets. *Inf Sci* 661:120183
12. Oktay O, Schlemper J, Folgoc L, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, Glocker B, Rueckert D (2018) Attention u-net: learning where to look for the pancreas. In: Medical imaging with deep learning

13. Peng D, Yu X, Peng W, Lu J (2021) Dgfau-net: Global feature attention upsampling network for medical image segmentation. *Neural Comput Appl* 33(18):12023–12037
14. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF (eds) *Medical image computing and computer-assisted intervention (MICCAI 2015)*. Springer International Publishing, Cham, pp 234–241
15. Salehi SSM, Erdogmus D, Gholipour A (2017) Tversky loss function for image segmentation using 3d fully convolutional deep networks. In: Wang Q, Shi Y, Suk H-I, Suzuki K (eds) *Machine learning in medical imaging*. Springer International Publishing, Cham, pp 379–387
16. Taghanaki SA, Zheng Y, Kevin Zhou S, Georgescu B, Sharma P, Xu D, Comaniciu D, Hamarneh G (2019) Combo loss: handling input and output imbalance in multi-organ segmentation, *Comput Med Imaging Graph* 75:24–33
17. Wu J, Xu H, Zhang S, Li X, Chen J, Zheng J, Gao Y, Tian Y, Liang Y, Ji R (2021) Joint segmentation and detection of COVID-19 via a sequential region generation network. *Pattern Recogn* 118:108006
18. Yap MH, Pons G, Martí J, Ganau S, Sentís M, Zwigelaar R, Davison AK, Martí R (2018) Automated breast ultrasound lesions detection using convolutional neural networks, *IEEE J Biomed Health Inf* 22:1218–1226

# Stirring up Biomarker Discovery for Cardiovascular Diseases Diagnosis: The FDR-RFE Pipeline



Harsh Bhasin , Chandsi, and Kapila Kumar

**Abstract** Cardiovascular diseases are disorders of heart and blood vessels and have become a cause of concern in the recent years. The identification of biomarkers in these diseases will help the clinicians to better manage and prevent the diseases. However, the biomarker identification in cardiovascular diseases dataset becomes challenging owing to limited samples and a large number of features. The conventional feature selection approaches do not perform well owing to the curse of dimensionality. This work proposes an ensemble feature selection approach combining a filter and a wrapper method and finding the common features from the two. The proposed work results in efficient model resulting in an accuracy of 97.01. The results are based on an extensive study employing hyperparameter tuning and focusing on the explainability of the model. It may be noted the proposed pipeline handles variance and bias gracefully. The work finds the important biomarkers for the disease and has the potential of assisting the clinicians in effective and efficient diagnosis of the disease.

**Keywords** Cardiovascular diseases · Feature selection · Machine learning · Wrapper methods · Biomarkers

## 1 Introduction

According to World Health Organization (WHO), cardiovascular diseases (CVDs) are a group of disorders of the heart and blood vessels and include coronary heart disease, cerebrovascular disease, rheumatic heart disease and other conditions [1]. Out of these, the heart disease encompasses a wide range of cardiovascular problems. It is important to diagnose the problem at its early stage to help the clinicians handle

---

H. Bhasin · Chandsi · K. Kumar ()

Manav Rachna International Institute of Research and Studies, Faridabad, Haryana, India  
e-mail: [kapila.set@mriu.edu.in](mailto:kapila.set@mriu.edu.in)

H. Bhasin

School of Computer Science Engineering & Technology, Bennett University, Greater Noida, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2025

A. Kumar et al. (eds.), *Proceedings of Fourth International Conference on Computing and Communication Networks*, Lecture Notes in Networks and Systems 1292,

[https://doi.org/10.1007/978-981-96-3250-3\\_2](https://doi.org/10.1007/978-981-96-3250-3_2)

the symptoms. The conventional methods are time-consuming, expensive and problematic to the patients. Machine learning (ML) helps us to efficiently and effectively handle this problem, provided sufficient amount of data is provided.

Artificial intelligence (AI) may be defined as the study and the development of machines that can behave and act like human beings or can behave and act rationally [2]. As per Tom Mitchell, ML is a subset of AI that can be defined as program that improves its performance with experience on a particular task [3]. It may be classified as supervised, unsupervised and semi-supervised. The supervised learning models learn from a teacher via a labelled dataset, whereas the unsupervised learning model learns patterns from a given dataset without a teacher. Tasks like classification and regression comes under the category of supervised learning, whereas clustering, etc., come under unsupervised learning. A machine learning pipeline includes pre-processing of data, extracting features and selecting features. Feature selection can be done using filter methods, wrapper methods, or heuristic methods [4].

ML is being used in exploring genomic data efficiently and effectively and finding out the patterns associated with a particular disease. According to the literature review, it helps in the classification of phenotypes and in analysing the prognosis of the disease. Scientists have been able to leverage the features from genomic data clubbed with clinical data to predict CVD and to identify biomarkers associated with a particular disorder [5].

In general, ML is being used for precision medicine that can be facilitated by finding out the significant biomarkers by using feature selection and feature extraction techniques. The aim is to use biomarker for the prediction of CVD in early stages and to reduce the number of features in the given dataset. This work uses the data obtained from [5]. It uses an ensemble of feature selection techniques to find the most pertinent features responsible for cardiovascular diseases (CVDs). The dataset we used had two major problems: (a) less number of samples and huge number of features and (b) imbalanced dataset. The existing works were not able to reduce the number of features as much, which motivated this study. The proposed work reduces the number of features to 34 from 751, using a novel feature selection method, which eventually decrease the computation cost and increase the accuracy.

This work is organized as follows. Section 2 presents an overview of materials used in this work and the proposed work. Section 3 presents the results and Sect. 4 concludes.

## 2 Materials and Methods

The dataset used in this work is obtained from [5] and consists of comprehensive transcriptomic data from individuals with cardiovascular disease and healthy individuals. The dataset includes gene expression profiles and clinical characteristics of these individuals, making it AI/ML ready for analysis. The data consists of 72 samples and 751 features. There are two classes to predict, namely ‘Control’ and ‘Case’ [5].

## 2.1 Proposed Work

The aim of this work is to find out the most pertinent features on which the type of CVD depends. The data was obtained from DeGroat et al. [5]. The data has 751 features and the label consisting of values 1 and 0. It is an imbalanced dataset consisting of 10 samples belong to the first class and 62 belong to the second class. As far as ML is concerned, the major challenges in the classification of this dataset are curse of dimensionality and imbalanced dataset.

The proposed work aims to develop a model that gracefully handles these problems. For finding out the most pertinent features, we used a combination of filter and wrapper methods. First of all, we found the Fisher Discriminant Ratio (FDR) [6] value of each feature using the following formula.

$$\text{FDR}(i) = \frac{(m_1(i) - m_2(i))^2}{(s_1^2(i) + s_2^2(i))}, \quad (1)$$

where  $m_1(i)$  is the mean of the samples of  $i$  in the features belonging to the class 0,  $m_2(i)$  is the mean of the samples of  $i$  in features belonging to the class 1,  $s_1(i)$  is the standard deviation of the samples of  $i$  in features belonging to the class 0, and  $s_2(i)$  is the standard deviation of the samples of  $i$  in features belonging to the class 1.

This was followed by arranging the FDR values in decreasing order and carrying out forward feature selection. We select minimum number of features for which the accuracy in the so obtained graph is maximized.

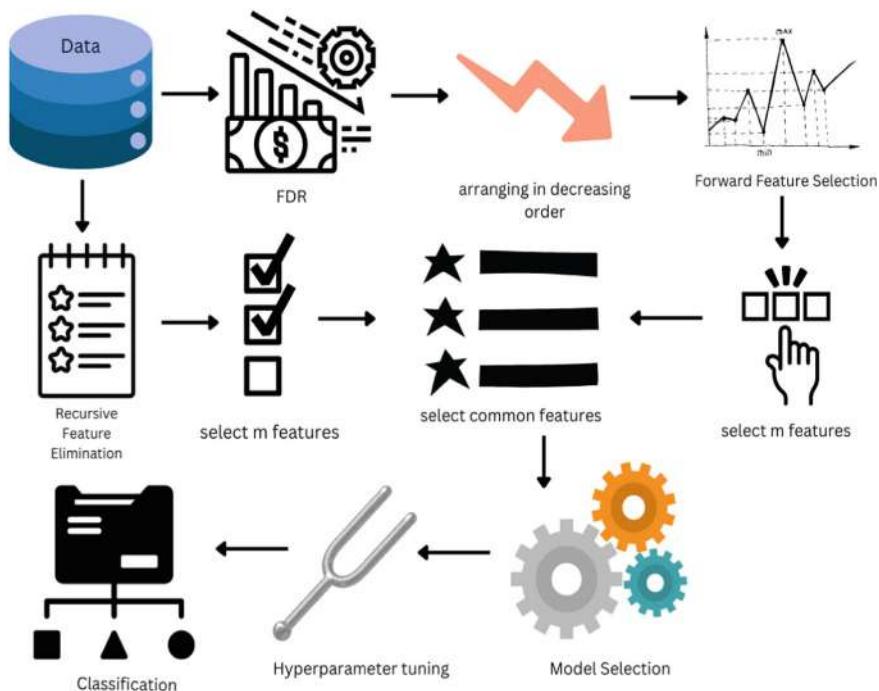
After doing this, we applied recursive feature elimination (RFE) using support vector machine (SVM) to find the top features for which the accuracy is maximum. From the above two sets, we find out the common features, and using those common features, we carried out classification. The selection of appropriate model and hyper-parameter tuning leads us to optimized model. The proposed model is shown in Fig. 1.

## 3 Experiments and Results

The dataset was loaded in the Google Drive, and experiments were carried out using Google Colab. The FDR was found, and forward feature selection was carried out. The results are shown in Fig. 2.

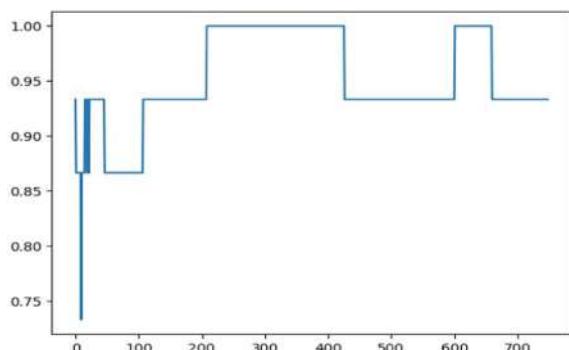
It was found that best results were obtained with 208 features. This was followed by applying RFE with support vector classifier having linear kernel. We extracted the selected features and found out the common features using the following formula.

```
common_features = set(selected_features_1) and
                  set(selected_features_2)
```



**Fig. 1** Selecting pertinent features from CVD dataset

**Fig. 2** Variation of performance with the number of features in forward feature selection



where selected\_features\_1 were obtained using FDR and selected\_features\_2 were obtained using RFE.

It was found that 34 features were common which are as follows:

[‘ENSG00000075624’, ‘ENSG00000102879’, ‘ENSG00000105835’,  
 ‘ENSG00000112303’, ‘ENSG00000116741’, ‘ENSG00000119535’,  
 ‘ENSG00000122862’, ‘ENSG00000130066’, ‘ENSG00000132475’,  
 ‘ENSG00000135821’, ‘ENSG00000136167’, ‘ENSG00000142347’,

'ENSG00000143226', 'ENSG00000150991', 'ENSG00000159840',  
 'ENSG00000160255', 'ENSG00000160593', 'ENSG00000162511',  
 'ENSG00000162747', 'ENSG00000163041', 'ENSG00000163464',  
 'ENSG00000168961', 'ENSG00000171051', 'ENSG00000180871',  
 'ENSG00000184009', 'ENSG00000188404', 'ENSG00000189067',  
 'ENSG00000197746', 'ENSG00000204592', 'ENSG00000206082',  
 'ENSG00000206172', 'ENSG00000206503', 'ENSG00000234745',  
 'ENSG00000266422'].

Different models were created for classifying the dataset using only these features. The results of the experiments are presented in Table 1.

It was observed that MLP classifier with hidden layer consisting of 100 neurons model gave the best results.

The proposed model gives the better results compared to the state of the art. DeGroat et al. [5] used statistical algorithms to find the most important biomarkers to classify patients of CVD and healthy controls. They used Pearson correlation, chi-square and analysis of variance (ANOVA) to find the most important features. They applied RFE but as per the paper due to high complexity, they were not able to obtain favourable results.

As an extension to this study, Baghdadi et al. [7] applied various classifiers to accomplish this task. However, due to absence of appropriate feature selection methods their performance was not upto the mark. This is the reason for carrying out this study.

The results obtained by various researchers for classifying CVD are shown in Table 2.

The results are also compared with the proposed work. Table 2 shows the model, accuracy and F1-measures obtained on the CVD dataset. Saha et al. [8] carried out the comprehensive review of various approaches for detection of CVD. The author's analysed various ML and DL techniques to diagnose the above. Liu et al. [9] used multimodal data and DL approaches for carrying out CVD diagnosis. The author's used modalities like X-ray and ECG to accomplish this task. Siddiqui et al. [10] used myocardial infarction diagnosis to accomplish the above task. Adam et al. [11]

**Table 1** Results

Classifier	Kernel/layers	Hidden layer (In case of NN)	Accuracy
SVM classifier	Kernel = linear		93.4
SVM classifier	Kernel = rbf		90.12
<b>MLP classifier</b>	<b>1 hidden layer</b>	<b>100 neurons</b>	<b>97.01</b>
MLP classifier	1 hidden layer	32 neurons	94.0
MLP classifier	1 hidden layer	16 neurons	93.2
MLP classifier	1 hidden layer	8 neurons	88.1
MLP classifier	2 hidden layers	(32,16) neurons	93.0
MLP classifier	2 hidden layers	(16,8) neurons	93.0

**Table 2** Comparison of the results with the existing models

Research work	Classifier	Accuracy	F1-score
Baghdadi et al. [7]	Linear discriminant	0.8696	0.8868
	Support vector machine	0.8841	0.9030
	K-neighbours	0.8841	0.9018
	AdaBoost	0.8659	0.8818
	Random forest	0.8877	0.9034
<b>Proposed Work</b>	<b>FDR_RFE_feature_selection</b>	<b>0.9701</b>	<b>0.9503</b>

found the biomarkers that can be used to diagnose the disease with greater accuracy and efficiency. Except for the above, this work was also compared with other works [12–20] and was found to be better than others in terms of accuracy of the number of features finally selected by the model.

## 4 Discussion and Conclusion

This paper carried out an extensive exploratory data analysis of the CVD dataset and found the most important features using a combination of FDR and wrapper methods. The model so-created uses only 34 features and is computationally effective and affective. An accuracy of 97.01 was obtained by the model. The work paves way for the use of ML in selecting biomarkers for CVD. The work is now being applied to different datasets to explore the generalizability of the model. The applicability of newer heuristic techniques like diploid and triploid genetic algorithms for feature selection is also being explored.

## References

1. Cardiovascular diseases (2019). [https://www.who.int/health-topics/cardiovascular-diseases#tab=tab\\_1](https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1)
2. Russell S, Norvig P (2009) Artificial intelligence: a modern approach, 3rd edn. Pearson
3. Mitchell T (1997) Machine learning. McGraw-Hill Professional
4. Alpaydin E (2016) Machine learning. MIT Press, London, England
5. DeGroat W, Abdelhalim H, Patel K et al (2024) Discovering biomarkers associated and predicting cardiovascular disease with high accuracy using a novel nexus of machine learning techniques for precision medicine. Sci Rep 14
6. Duda RO, Hart PE (1973) Pattern classification and scene analysis. Wiley
7. Baghdadi NA, Farghaly Abdelalim SM, Malki A et al (2023) Advanced machine learning techniques for cardiovascular disease early detection and diagnosis. J Big Data 10
8. Saha P, De A, Roy SD, Bhowmik MK (2024) A comprehensive review on deep cardiovascular disease detection approaches: its datasets, image modalities and methods. Multimedia Tools Appl

9. Liu Y, Li D, Zhao J, Liang Y (2023) Enhancing heart failure diagnosis through multi-modal data integration and deep learning. *Multimedia Tools Appl* 83:55259–55281
10. Siddiqui HUR, Zafar K, Saleem AA et al (2023) Artificial intelligence-based myocardial infarction diagnosis: a comprehensive review of modern techniques. *Multimedia Tools Appl* 83:41951–41979
11. Adam CA, Şalaru DL, Prisacariu C et al (2022) Novel biomarkers of atherosclerotic vascular disease—latest insights in the research field. *IJMS* 23:4998
12. Annabel LSP, Sruthi BS, Rohini M, Svetna BS (2024) Effective prediction of cardiovascular disease using deep learning. In: ICT: applications and social interfaces, pp 259–270
13. Jain PK, Tadepalli KV, Roy S, Sharma N (2023) Exploring deep learning for carotid artery plaque segmentation: atherosclerosis to cardiovascular risk biomarkers. *Multimedia Tools Appl* 83:42765–42797
14. Nguyen H-H, Vo T-T (2024) Enhancing cardiovascular health monitoring through IoT and deep learning technologies. *SN Comput Sci* 5
15. Parashar G, Chaudhary A, Pandey D (2024) Machine learning for prediction of cardiovascular disease and respiratory disease: a review. *SN Comput Sci* 5
16. Dhiyanesh B, Ammal SG, Saranya K, Narayana KE (2024) Advanced cloud-based prediction models for cardiovascular disease: integrating machine learning and feature selection techniques. *SN Comput Sci* 5
17. Pandey A, Shivaji BA, Acharya M, Mohbey KK (2024) Mitigating class imbalance in heart disease detection with machine learning. *Multimedia Tools Appl*
18. Rimal Y, Paudel S, Sharma N, Alsadoon A (2023) Machine learning model matters its accuracy: a comparative study of ensemble learning and AutoML using heart disease prediction. *Multimedia Tools Appl* 83:35025–35042
19. Türk F (2024) Investigation of machine learning algorithms on heart disease through dominant feature detection and feature selection. *SIViP* 18:3943–3955
20. Rimal Y, Sharma N (2023) Hyperparameter optimization: a comparative machine learning model analysis for enhanced heart disease prediction accuracy. *Multimedia Tools Appl* 83:55091–55107

# Support Vector Machines and Slime Mold Optimization Algorithms for SQL Injection Detection



Zainab H. Al-Araji and Hasanen Alyasiri

**Abstract** The web is a critical application in a variety of online services, including e-banking and e-commerce. However, the potential threat of SQL Injection (SQLI) attacks to sensitive data that is processed online is on the rise. SQLI can be employed to compromise confidential information from the targeted database and obtain unauthorized access by exploiting vulnerabilities in online applications. This paper introduces a detection system that is powered by machine learning and is designed to mitigate such attacks. A hybrid system that integrates Slime Mold Algorithms (SMA) and Support Vector Machine (SVM) was proposed as a novel method for safeguarding web applications from SQLI attacks. The SMA algorithm is employed in this paper to determine the most effective values for SVM parameters to enhance performance. To evaluate the proposed system performance, we compiled a dataset including normal and attack SQL queries In addition to that Extracted a set of statistical features that perfectly describe each query. The experimental results show that the proposed method SMA-SVM obtained 0.9931 concluding that this method effectively detects SQL injection attacks rather than compared to single-SVM.

**Keywords** Support vector machine · SQL injection attack · Bio-inspired algorithm · Slime mold algorithm

## 1 Introduction

In today's interconnected world, web applications are becoming more and more popular in the industry to provide cost-effective and ubiquitous remote services [1]. Companies and organizations depend on web applications to provide services to

---

Z. H. Al-Araji ( ) · H. Alyasiri

Department of Computer Science, College of Computer Science and Mathematics, University of Kufa, Najaf, Iraq

e-mail: [zainabh.alfaham@student.uokufa.edu.iq](mailto:zainabh.alfaham@student.uokufa.edu.iq)

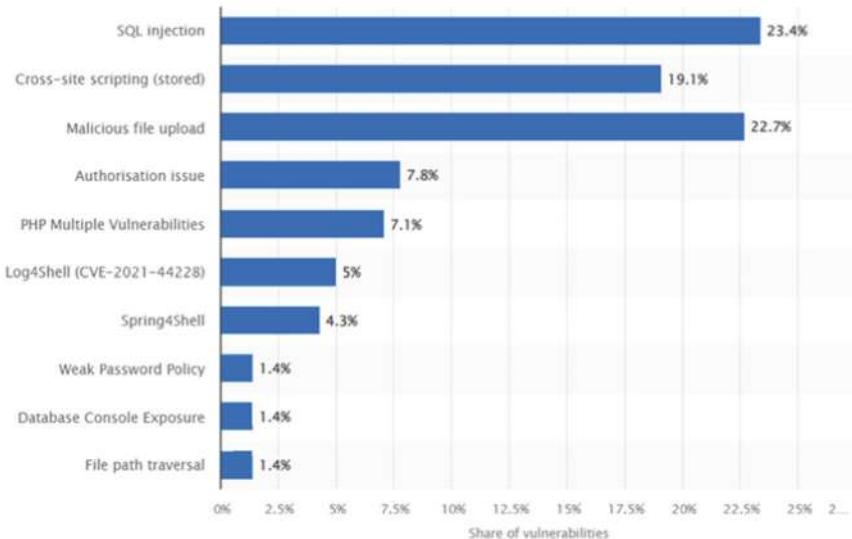
H. Alyasiri

e-mail: [hasanen.alyasiri@uokufa.edu.iq](mailto:hasanen.alyasiri@uokufa.edu.iq)

clients considering that the infrastructure is internet-based, lowering the cost to the absolute minimum e-banking, shopping, e-governance, and bookings are examples of services whose speed and utility flexibility boost daily productivity. The global web development market was \$5.6B in 2021 and is projected to reach \$130.97B by 2032 [2]. Nevertheless, as they have grown in popularity so have the rate of attacks and threat complexity. The 2024 Global Threat Analysis Report reveals that there is a 171% growth in total malicious web application and API transactions in 2023 [3]. Hence, security teams should deal with various threats and ensure a robust response to new security challenges.

SQL injection (SQLI) attack is a common attack vector that employs malicious Structured Query Language (SQL) code to access and manipulate databases which leads to gaining unauthorized access and compromising confidential information [4, 5]. According to a recent report, SQLI attacks appeared as the primary source of critical vulnerabilities in web applications globally, with 23% (see Fig. 1) [6]. In 2013, 2017, and 2021 it ranked one of the ten most often occurring web application vulnerabilities sequentially [7]. In some circumstances, an attacker can run a denial-of-service attack or employ a SQLI assault to compromise the underlying server or another back-end infrastructure. To handle this type of assault, web developers should thus include security elements including input validation, output encryption, and robust authentication systems, and routinely change their online programs. Still, conventional defensive technologies are insufficient to counter such threats [8]. Thus, it is necessary to implement fresh approaches to support defense systems. In various spheres of computer science, including cybersecurity, optical character recognition, and product recommendation, machine learning (ML) has shown success [9–11]. Security teams have so been aggressively searching for ML techniques to overcome the constraints in precisely identifying and stopping different types of attacks.

A commonly used ML algorithm is the Support Vector Machine (SVM), which has been extensively applied in both academia and industry. Despite its widespread use, the performance of SVM-based systems is highly dependent on the model's parameters. There are many parameters in the SVM model, hence choosing the best values using human knowledge might be difficult and demanding work [12]. Our work presents a hybrid technique combining the SVM classifier with the Slim Mold Algorithm (SMA) which is a type of Bio-inspired algorithm. SMA has recently received much attention from researchers because of its simple structure, excellent optimization capabilities, and acceptable convergence in dealing with various types of complex real-world problems to handle these difficulties [13]. Hence, the issue of parameter adjustment, a concern we want to address with the SMA method, typically limits the great performance of SVM. In our suggested approach, the accuracy and resilience of the SVM classifier are applied to classification problems. Conversely, the SMA is used to adjust SVM model parameters to guarantee the best performance. This hybrid strategy makes use of the advantages of both methods: the effective parameter optimization capacity of SMA and the great accuracy of SVM. Our work attempts to create a more dependable and effective assault detection system by combining these two techniques. This system increases threat identification's accuracy while lowering parameter tuning's time and effort required.



**Fig. 1** Common web application risks as of 2023 [6]

We compiled 47,464 SQL searches for this proposed study, of which 21,664 were malicious SQL searches to test and develop our model. As such, this paper mostly contributes in the following ways:

- Proposed a SQLI detection system based on Support Vector Machine and Slim Mold Algorithm. Our experiments show that utilizing the SMA optimization technique to choose smoothing parameters improves SVM performance. This indicates thus that the suggested approach can successfully identify SQLI attacks.
- We gathered the most current SQL queries and then generated an efficient 18 numerical feature set that characterizes every query to train the proposed approach utilizing a vast dataset of both regular and malicious.
- We investigate non-security properties of the produced detector such as processing time.

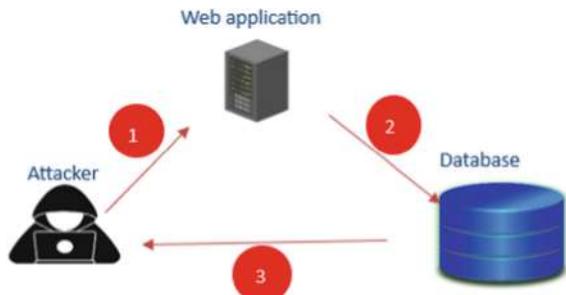
The paper is outlined in detail below. Section 2 describes SQLI Attack. Section 3 details the algorithms employed in this investigation, dataset description, feature extraction, and performance measures. Section 4 illustrates achieved results and compares them with state-of-the-art studies published for SQL injection detection, while Sect. 5 discusses and Limitations. Section 6 provides the paper's conclusion and future directions.

## 2 Understanding SQL Injection Attack

Attackers undertake substantial risks, such as using SQLI attacks, to pilfer confidential information, compromise individuals' identities, and illicitly obtain access to sensitive data. Security risks encompass many actions such as unauthorized disclosure of information, manipulation of data, unauthorized access to someone else's account, execution of malicious code, and disruption of services [14]. SQLI attacks exploit some of the most perilous vulnerabilities to cause harm. SQLI attacks enable malicious individuals to gain unauthorized access to and distribute sensitive information such as login details, passwords, credit card numbers, and personal data. Adversaries can exploit this information to steal identities, perpetrate fraud, and execute other nefarious endeavors. In addition, they possess the capability to modify, erase, or seize control of all the information stored in the database. Consequently, there is a risk of losing crucial data and causing harm to your business's reputation and financial stability. Adversaries can install malicious software on the targeted computer, enabling them to remotely access and execute it. This enables unauthorized parties to access, modify, or manipulate data. SQLI attacks can also be employed for conducting denial-of-service assaults by overwhelming the targeted system with an excessive number of SQL requests, resulting in system failure or cessation of operation. This detrimental outcome negatively impacts the system's reputation and incurs financial expenses. The user's text is [15] Companies can mitigate SQLI attacks by adhering to security best practices, conducting periodic risk assessments, and employing sophisticated tools, such as machine learning-based methods, to detect and prevent such attacks (refer to Fig. 2). The three categories of SQLI [16] are as follows:

1. Error-Based: “1’UNION select 1, 2, 3+” → gives an error if the tables have a smaller number of columns than mentioned in the query.
2. Union Query SQLI: “1’UNION select database (), version ()+” → name of the database and version are returned. Any information can be retrieved by mentioning it in the query.

**Fig. 2** SQL injection attack workflow



3. Blind SQLI: Attackers send a query asking for either True or False; this information helps one deduce the type of data kept.
- b. Time-based: Attackers mark whether the outcome is True or False depending on the time the database responds.

### 3 Suggested Approach

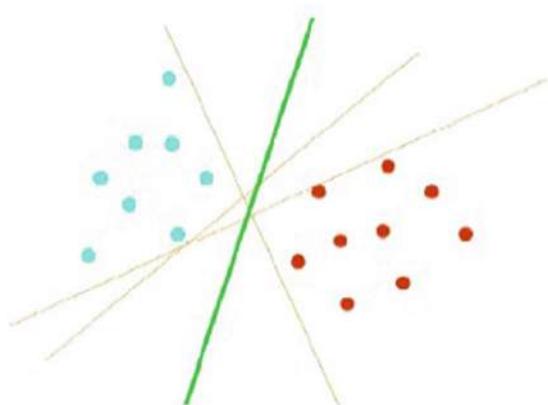
The support vector machine (SVM) classifier is used in the proposed hybrid approach to attack/normal categorization. Slime Mold Algorithm (SMA) has been used to optimize the fundamental parameters of SVM to get the best possible categorization accuracy.

#### 3.1 Support Vector Machine (SVM)

The term SVM [17] is commonly used to refer to classification using support vector methods, whereas SVM is used to explain regression using support vector methods. SVM is a valuable method for categorizing data. The classification problem can be narrowed down to the two-class problem without sacrificing generality. The objective of this issue is to segregate the two classes using a function that is derived from the given instances. The objective is to develop a classifier that has strong performance in unfamiliar instances, demonstrating good generalization as shown in Fig. 3. Numerous linear classifiers can separate the data, but only one of them maximizes the margin, which is the distance between it and the nearest data point of each class. The linear classifier is referred to as the optimal separating hyperplane [18].

Usually involving certain data instances, a classification task consists of training and evaluating data. Every occurrence in the training set consists of one “target value”

**Fig. 3** The perfect separating HyperPlane



(class labels) and numerous “attributes” (features). SVM aims to generate a model predicting the target value of data instances in the testing set given just the attributes.

1. Linear:  $K(x_i, x_j) = x_i^T \cdot x_j$
2. Polynomial: The Polynomial kernel of degree d is of the form.

$$K(x_i, x_j) = (x_i, x_j)$$

3. RBF: The Gaussian kernel, known also as the radial basis function, is of the form

$$K(X_i, X_j) = \exp\left(-\frac{(x_i, x_j)}{2\sigma^2}\right)$$

4. Sigmoid: The Sigmoid kernel is of the form

$$K(x_i, x_j) \tanh(k(x_i, x_j) + r)$$

By use of a nonlinear transformation, the RBF kernel maps samples onto a higher dimensional space. Unlike the linear kernel, this lets it manage scenarios when the link between class labels and characteristics is nonlinear. Furthermore, it can be shown that the linear kernel is a particular case of the RBF kernel. More precisely, it can be demonstrated that the RBF kernel with given parameters ( $C, r$ ) performs equivalently to the linear kernel with a penalty value of  $C$ . Moreover, for particular values, the sigmoid kernel shows corresponding behavior to the radial basis function (RBF). Our work has used Support Vector Machine (SVM) technology in a fresh way to find SQLI attacks. Analyzing the datasets of the original question and the suspect query helps one to classify dubious searches. After learning the dataset, the algorithm uses this knowledge depending on the learning method to categorize the searches. Through optimal learning strategies and rigorous design factor consideration, our system achieves accurate classification [19] (Table 1).

### **3.2 Slime Mold Algorithm (SMA)**

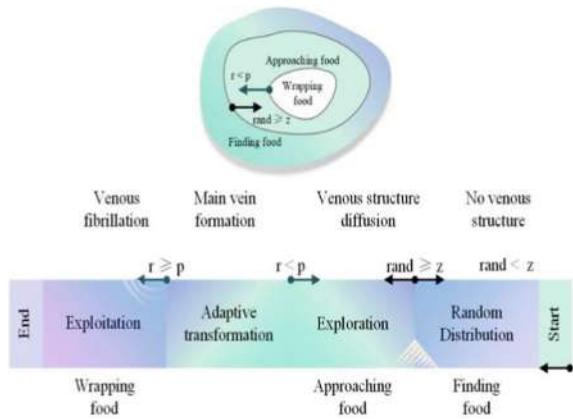
The SMA is a meta-heuristic method that operates on a population level and was recently introduced by [20]. The form of the SMA resembles that of the acellular slime mold *Physarum polycephalum* while quickly seeking sustenance. The term “Slime Mold” refers to a type of organism. The organism consists of an intricate network of linked tubes that transport cytoplasm throughout its structure. Slime Mold possesses this ability due to its unique anatomical structure (see Fig. 4), which enables it to create intricate networks of veins connecting its several sources of nourishment, so enabling it to consume all of them simultaneously. Once Slime Molds find a source of food, their biochemical oscillator triggers contraction waves that travel through their venous system, causing the cytoplasm to flow via tubular

**Table 1** Support vector machine parameters

Parameter	Range	Explanation
C-paras	(0.01, 10)	Change the margin and misclassification with the $C$ parameter. Low misclassification and a high $C$ value are opposite; a high $C$ value indicates higher misclassification but a high margin
Kernel-paras	“linear”, “poly”, “rbf”, and “sigmoid”	Hyperplane calculations are done via kernel parameters. Linear option can be applied if data can be separated with a linear line. That one is the easiest to compute hyperplane. But rbf or poly can be applied to transform high dimensions if data cannot be separated with a linear line. (It is called Kernel Trick as well). Rbf is generally preferable to poly
Degree-value	(1, -5)	The degree parameter is the Polynomial kernel degree. Other kernels all disregard it. A higher degree corresponds to a more adaptable hyperplane
Coef0-value	(-1, 1)	Utilized to the sigmoid and polynomial kernel functions of the SVM algorithm
Gamma-value	(0.1–0.9)	Gamma is about nonlinear kernels and attempts to provide greater fitting; nevertheless, overfitting might result from a too-high gamma parameter selection
Random-state	None	Regulates the pseudo-random number generating to shuffle the data for probability estimations. Neglected when the likelihood is False. Pass an int for repeatable output over several function calls

veins. The velocity of cytoplasmic streaming is correlated with the thickness of the vein’s wall. Consequently, when the rate of cytoplasmic streaming rises, the vein becomes thicker; conversely, when the rate drops, the vein becomes thinner. The Slime Mold utilizes both positive and negative stimuli to navigate toward sources of nourishment [21].

**Fig. 4** The schematic diagram of the SMA [22]

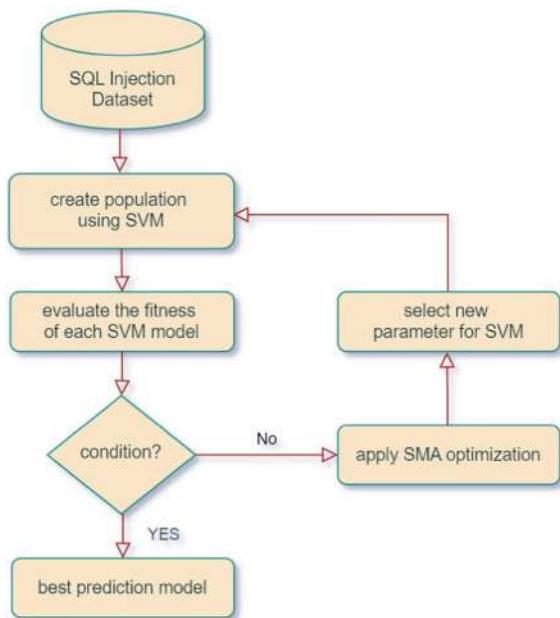


### 3.3 Optimization of SVM Using SMA

We use the proposed Hybrid approach utilizing a Slim Mold Algorithm (SMA) to optimize the parameters of the Support Vector Machine (SVM) algorithm. The objective is to identify SVM parameters that enhance the classification accuracy of SQL queries as illustrated in Fig. 5. This hybrid model's fitness criterion is based on accuracy. The optimization of SVM using the Slim Mold Algorithm (SMA) involves iteratively adjusting the parameters of SVM models to enhance their performance in detecting SQLI attacks. Initially, a population of SVM models is created, each with different parameter settings. The fitness of these models is evaluated based on performance metrics such as accuracy. If the stopping condition (e.g., maximum iterations or performance threshold) is not met, SMA is applied to explore and exploit the search space effectively. SMA is used to generate the best solutions, optimizing the SVM parameters. This process iterates, with the new parameters leading to a new set of SVM models, until the best model is found that meets the desired performance criteria. This optimized model is then selected as the final prediction model for detecting SQLI attacks.

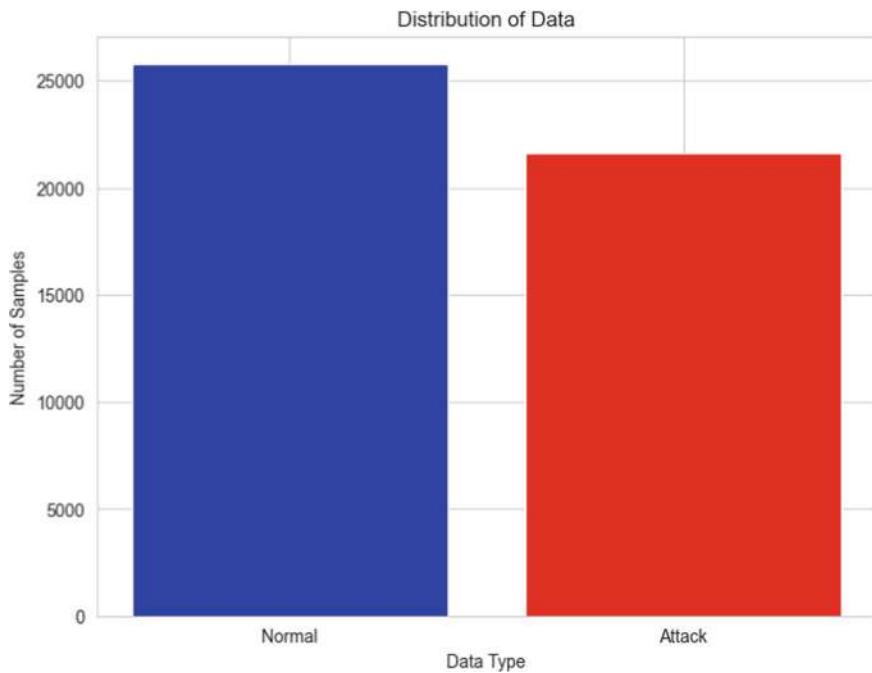
We utilized Python to apply the SVM algorithm using the sci-kit-learn library [23]. Additionally, the SMA algorithm is available through the Mealpy package [24] which is the largest Python library in the world for most of the cutting-edge meta-heuristic algorithms (nature-inspired algorithms, black-box optimization, global search optimizers). For future comparison, the key parameters set in the SMA algorithm are (**epoch** represents the maximum number of iterations, **Pop-size** specifies the size of the population, and **p-t** indicates the probability threshold). The remaining parameters are automatically determined by the library.

**Fig. 5** The overall architecture of the proposed technique



### 3.4 Dataset Description

The carefully built dataset presented in this paper is intended to be used in the training of supervised machine learning algorithms that identify SQLi threats. We manually gathered datasets from Kaggle [25] and GitHub [26, 27]. There are 47,464 distinct SQL queries in it, including both legitimate and malicious ones. All of the components related to SQL queries are contained in each entry of this dataset, including semicolons, single quotes, intermediate data, text fragments, and SQL keywords. Each row in the dataset has a binary label, where attack SQL queries are indicated by 1 and normal queries by 0. The dataset is built with 25,800 benign queries and 21,664 destructive queries (see Fig. 6). The single-column display of this binary labeling facilitates the identification of the kind of query. This work's second primary contribution is the creation of a 19-feature numeric training dataset. Through feature homogeneity across the dataset, this study aims to raise machine learning algorithm accuracy and precision. 18 useful numerical features were extracted from typical SQLi datasets as the first step in the development process. The source code of every query in the chosen original dataset was used to create these features. The design of the dataset consists of one dependent feature, which acts as the label designating the type of query (0 for normal and 1 for malicious), and eighteen independent characteristics. Constants, punctuation, logical operators, the duration of the question, and the number of nested queries are among the purely numerical data that are extracted. As a result, there are 47,464 records in the improved dataset, and each record has 18 extracted attributes. This methodical approach to feature homogeneity greatly



**Fig. 6** Dataset distribution

**Table 2** Sample of normal data

---

Statement
1 select * from user where age > 20
2 insert into manager (Index, Price) values (11, 90)
3 delete from products where size = "122"

---

**Table 3** Sample of attack data

---

Statement
1 select * from name where 1 = 1 and "s" = "s"
2 selct* from data where index = "or 1 = 1#" and pw = md5("")

---

improves the effectiveness of machine learning algorithms used to identify SQLI threats (Tables 2 and 3).

**Table 4** Description of extracted features in the dataset

Features	Overview
1	Total number of single quotations in a query (‘’)
2	Total number of double quotations in a query (“”)
3	Total number of punctuations in a query “;”, “.”, “( )”, “[ ]”, “@”, “!”, “*”, “/”, “?”, “,”, “_”, “_2”, “_”, “...”, “@”
4	Total number of white spaces in a query
5	Total number of normal keywords in a query (“SELECT”, “OR”, “INSERT”, “AND”, “IN”, “JOIN”, “FROM”, “UNION”, “NOT”, “WHERE”)
6	Total number of percentage (%) symbols in a query
7	Total number of logical operators in a query ( )
8	Total number of operators in a query (+, -, *, /, <, >)
9	Total number of null values in a query
10	Total number of letters in a query (a-z) (A-Z)
11	Total number of digits in a query (0-9)
12	Total number of special characters in a query (#, \$,!.)
13	Entropy
14	Total number of the OR
15	Total number of the AND
16	Number of comments (-)
17	Length
18	Number of semicolon ( ; )

### 3.5 Feature Extraction

This section focuses on the identification of 18 crucial characteristics that accurately represent important elements of web traffic and behaviors. These features are necessary for instructing our proposed approach to efficiently identify and counteract web threats (Table 4).

## 4 Results

We used a K-fold cross-validation method to lessen dependency on the random seed and estimate biases in performance evaluation. In Table 5, results of the SMA-SVM and SVM approach in terms of Accuracy, Precision, DR, and F1-score from each fold were used. Additionally, we report the average plus the Standard Deviation (STD). The results show that SMA-SVM performs better than single-SVM for each fold.

In [28], the algorithms (KNN, RF, XGB, and LSTM) used for SQLI detection and achieved execution times ranging from 0.967 to 133.7 s, while our algorithm

**Table 5** Results achieved using SVM and SMA-SVM

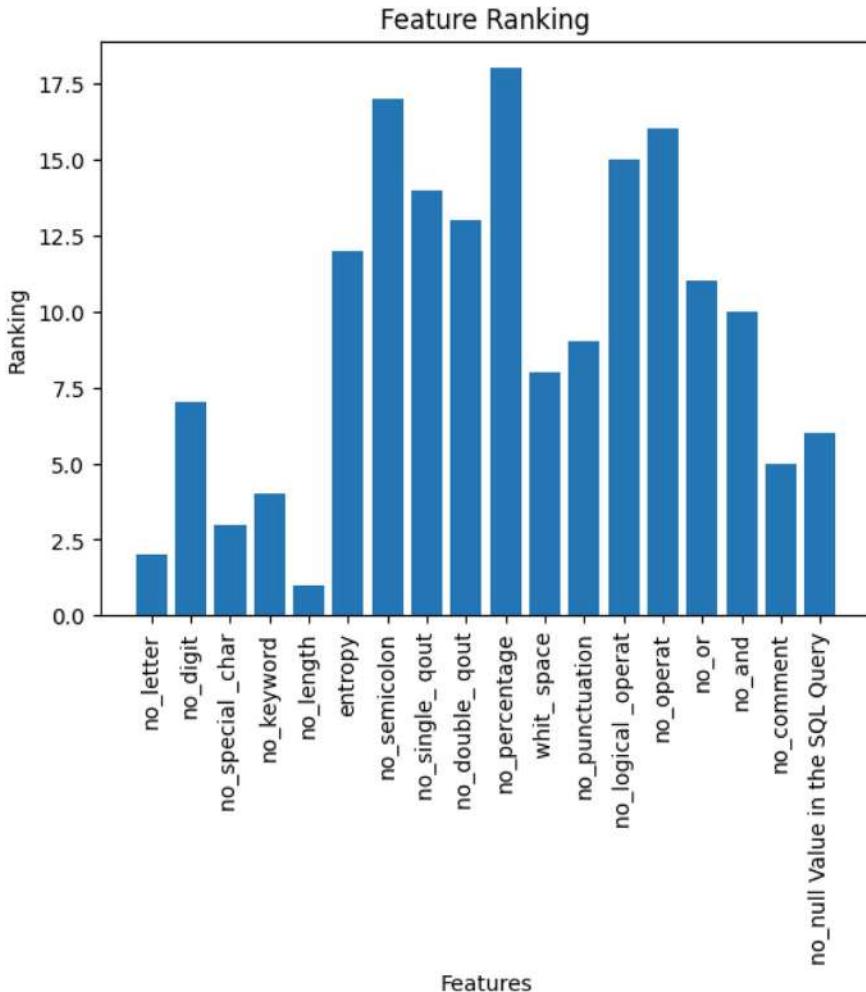
Fold	SVM				SMA-SVM			
	Accuracy	Precision	DR	F1-score	Accuracy	Precision	DR	F1-score
1	0.975	0.975	0.974	0.975	0.994	0.996	0.992	0.994
2	0.975	0.976	0.974	0.975	0.995	0.995	0.993	0.994
3	0.977	0.978	0.977	0.977	0.986	0.986	0.982	0.984
4	0.974	0.974	0.973	0.974	0.994	0.994	0.994	0.994
5	0.977	0.978	0.976	0.977	0.995	0.996	0.992	0.994
AVR	0.976	0.976	0.975	0.975	0.993	0.993	0.991	0.992
STD	0.0014	0.001	0.001	0.001	0.0040	0.0040	0.004	0.004

**Fig. 7** Barplots for testing time achieved by SMA-SVM

achieved an average execution time of 0.293 s, which shows a significant performance improvement compared to the other algorithms used in the study as shown in Fig. 7.

We employed a Support Vector Machine (SVM) to determine the ranks of features to detect SQLI attacks. At first, we collected diverse attributes of SQL queries, such as the count of letters, digits, special characters, keywords, ...). Next, we employed a feature ranking technique based on SVM. This involved training an SVM model on the dataset and examining the weights assigned to each feature. SQL queries that had greater weights for certain features were considered more influential in differentiating between legitimate and fraudulent queries. Figure 8 displays the results, with each bar representing a feature and its height indicating its significance in the classification task. By prioritizing the most highly ranked features, we improved our model by making it more understandable and computationally efficient. This method led to the development of a stronger and more efficient SVM classifier for identifying SQLI attacks.

This bar chart illustrates that the features “no\_semicolon” and “no\_percentage” possess the highest ranks, suggesting that these are the most crucial factors for



**Fig. 8** Feature ranking

detecting SQL injection. These traits are expected to capture fundamental attributes that have a high correlation with fraudulent searches. Conversely, the features “no\\_letter” and “length” have the lowest rank, indicating that they have a minor contribution to the detection process and are less useful in discriminating between regular and abnormal SQL queries.

We have compared our novel proposed approach’s performance with state-of-the-art studies published for SQL injection detection. We have used recent studies for comparison. The analysis reveals that our proposed approach outperformed the state-of-the-art approach with a high accuracy performance of 0.976 when using SVM without optimization and achieved similar in some cases and better performance

**Table 6** Comparison with other works on SQLI attacks detection

References	Algorithm	Dataset Size	Year	Accuracy	Precision	DR	F1-score
Maha et al. [29]	RNN	30,907	2023	94%	95%	90%	92%
Ketema et al. [30]	CNN	4199	2022	97%	NA	NA	NA
Jaradat et al. [31]	GA + GB	2300	2023	0.95	0.94	0.94	0.94
Venkatramulu et al. [32]	CSA	7539	2018	0.937	0.883	NA	NA
This study	SVM	47,464	2024	0.976	0.976	0.977	0.977
This study	SMA-SVM	47,464	2024	0.993	0.993	0.991	0.992

overall When applying the hybrid approach SMA-SVM It achieved accuracy up to 0.993 (see Table 6). Additionally, the dataset used in our testing was larger than the compared studies.

## 5 Discussion and Limitations

Based on performance measures, the suggested hybrid system integrating Support Vector Machine (SVM) and slime mold algorithm (SMA) presents a novel technique to identify SQLI attacks with impressive accuracy. Using parameter optimization applying SMA, the study shows notable enhancements over conventional SVM models, attaining an accuracy of 0.9931, which is higher than several state-of-the-modern techniques. The study had constraints, though. Computationally taxing is the reliance on ideal parameter tweaking for SVM even with SMA optimization. Although large, the dataset might not include all conceivable SQLI variants, therefore affecting the generalizability of the model. Further validation comes from the hybrid system's performance in practical applications with various and challenging searches.

## 6 Conclusion

SQLI attack is a web application safety threat whose technique is based on injecting malicious code into user input fields. An attacker's objective is to target database information that can be affected according to the injection type. the attacker may gain unauthorized access or manipulate database information. This study presents a new approach to protect web applications from SQLI attacks using a hybrid approach that combines Support Vector Machine (SVM) and Slime Mold Algorithms (SMA).

The SMA algorithm is used to optimize the parameters of the SVM model, allowing it to achieve higher performance in detecting SQLI attacks. Experimental results reveal that the proposed SMA-SVM method has achieved promising results compared with recently published studies with an impressive accuracy score of 0.9931.

In future, we plan to use this model to detect a different form of cyber-attacks for instance cross-site scripting. Another avenue of research would be providing security using the proposed method for other environments such as IoT.

**Acknowledgements** The author would like to thank the anonymous reviewers for their efforts.

**Disclosure of Interests** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Rani GS, Sarika S, Rupa P (2021) A study of prevention and detection analysis of SQL injection attack. AIP Conf Proc 2358:0500152021 <https://doi.org/10.1063/5.0059318>
2. Web development market size, share, growth, and industry analysis, by type (websites, web apps), by application (large businesses, small businesses, colleges and universities, government, non-profits), regional insights, and forecast to 2032, Mar. 2024 (Online). Available: <https://www.businessresearchinsights.com/market-reports/web-development-market-109039>. Accessed 25 Jun 2024
3. Radware (2024) Global threat analysis report (Online). Available: <https://www.radware.com/threat-analysis-report-2024/>. Accessed 03 Jul 2024
4. Abdullayev V, Chauhan DAS (2023) SQL injection attack: quick view. Mesopotamian J CyberSecurity 2023:30–34. <https://doi.org/10.58496/mjcs/2023/006>
5. Muhammad T, Ghafory H (2022) SQL injection attack detection using machine learning algorithm. Mesopotamian J CyberSecurity 2022:5–17. <https://doi.org/10.58496/MJCS/2022/002>
6. Borgeaud A, Distribution of web application critical vulnerabilities worldwide as of 2023 (Online). Available: <https://www.statista.com/statistics/806081/worldwide-application-vulnerability-taxonomy/#statisticContainer>. Accessed 21 Jun 2024
7. The OWASP Top 10 2021 (Online). Available: <https://owasp.org/Top10/>. Accessed 03 Jul 2024
8. Khalaf O, Sokiyna M, Alotaibi Y, Alsufyani A, Alghamdi S (2021) Web attack detection using the input validation method: DPDA theory. Comput Mater Contin 68:3167–3184. <https://doi.org/10.32604/cmc.2021.016099>
9. Alyasiri H, Clark JA, Kudenko D (2018) Applying Cartesian genetic programming to evolve rules for intrusion detection system. In: IJCCI, pp 176–183
10. Sommer R, Paxson V (2010) Outside the closed world: on using machine learning for network intrusion detection. In: 2010 IEEE symposium on security and privacy. IEEE, pp 305–316
11. Jundi ZZ, Alyasiri H (2023) Android malware detection based on grammatical evaluation algorithm and XGBoost. In: 2023 Al-Sadiq International conference on communication and information technology (AICCIT). IEEE, pp 70–75
12. Hacham SAK, Uçan ON (2023) Detection of malicious SQL injections using SVM and KNN algorithms. In: ISAS 2023—7th international symposium on innovative approaches and smart technology proceedings, pp 1–5. <https://doi.org/10.1109/ISAS60782.2023.10391560>
13. Chen H, Li C, Mafarja M, Heidari AA, Chen Y, Cai Z (2023) Slime mould algorithm: a comprehensive review of recent variants and applications. Int J Syst Sci 54(1):204–235
14. Kumar A, Binu S (2018) Proposed method for SQL injection detection and its prevention. Int J Eng Technol 7(2.6):213. <https://doi.org/10.14419/ijet.v7i2.6.10569>

15. Khan JR, Farooqui SA, Siddiqui AA (2023) A survey on SQL injection attacks types & their prevention techniques. *J Indep Stud Res Comput* 21(2):10–13. <https://doi.org/10.31645/jisrc.23.21.2.1>
16. Pallam R, Konda SP, Manthripragada L, Noone RA (2021) Detection of web attacks using ensemble learning. *Learning* 3(4):5
17. Kim T, Pak W (2022) Robust network intrusion detection system based on machine-learning with early classification. *IEEE Access* 10:10754–10767
18. Rawat R, Kumar Shrivastav S (2012) SQL injection attack detection using SVM. *Int J Comput Appl* 42(13):1–4. <https://doi.org/10.5120/5749-7043>
19. Anupam S, Kar AK (2021) Phishing website detection using support vector machines and nature-inspired optimization algorithms. *Telecommun Syst* 76(1):17–32. <https://doi.org/10.1007/s11235-020-00739-w>
20. Bhandakkar AA, Mathew L (2023) Merging slime mould with whale optimization algorithm for optimal allocation of hybrid power flow controller in power system. *J Exp Theor Artif Intell* 35(7):973–1000. <https://doi.org/10.1080/0952813X.2022.2040598>
21. Gharehchopogh FS, Ucan A, Ibrikci T, Arasteh B, Isik G (2023) Slime mould algorithm: a comprehensive survey of its variants and applications, vol 30, no 4. Springer, Netherlands. <https://doi.org/10.1007/s11831-023-09883-3>
22. Chen H, Li C, Mafarja M, Heidari AA, Chen Y, Cai Z (2023) Slime mould algorithm: a comprehensive review of recent variants and applications. *Int J Syst Sci* 54(1):204–235. <https://doi.org/10.1080/00207721.2022.2153635>
23. Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2(3):1–40. <https://doi.org/10.1145/1961189.1961199>
24. Van Thieu N, Mirjalili S (2023) MEALPY: an open-source library for latest meta-heuristic algorithms in Python. *J Syst Archit* 139:102871
25. Vedant H, Dataset (Online). Available: <https://www.kaggle.com/datasets/henilvedant/sqlinjection-payload>. Accessed 01 Nov 2023
26. Taşdelen İ, Dataset (Online). Available: [https://github.com/payloadbox/sql-injection-payload-list/blob/master/Intruder/exploit/Auth\\_Bypass.txt](https://github.com/payloadbox/sql-injection-payload-list/blob/master/Intruder/exploit/Auth_Bypass.txt). Accessed 25 Oct 2023
27. Migolovanov, Dataset (Online). Available: <https://gist.github.com/migolovanov/432fe28c8c7e9fa675ab3903c5eda77f#file-libinjection-bypasses-txt>. Accessed 25 Oct 2023
28. Thalji N, Raza A, Islam MS, Samee NA, Jamjoom MM (2023) AE-Net: novel autoencoder-based deep features for SQL injection attack detection. *IEEE Access* 11(December):135507–135516. <https://doi.org/10.1109/ACCESS.2023.3337645>
29. Alghawazi M, Alghazzawi D, Alarifi S (2023) Deep learning architecture for detecting SQL injection attacks based on RNN autoencoder model. *Mathematics* 11(15):3286
30. Ketema A (2022) Developing SQL injection prevention model using deep learning technique. St. Mary's University
31. Jaradat AS (2023) Genetic optimization techniques for enhancing web attacks classification in machine learning. In: 2023 IEEE international conference on dependable, autonomic and secure computing. International conference on pervasive intelligence and computing. International conference on cloud big data computing. International conference on cyber science and technology congress, pp 130–136. <https://doi.org/10.1109/DASC/PiCom/CBDCom/Cy59711.2023.10361399>
32. Venkatramulu S, Guru Rao CV (2018) CSES: Cuckoo search based exploratory scale to defend input-type validation vulnerabilities of HTTP requests, vol 712. Springer, Singapore. [https://doi.org/10.1007/978-981-10-8228-3\\_23](https://doi.org/10.1007/978-981-10-8228-3_23)

# Sustainable Sensory Based Automated Water Regulation System in Energy Constrained Domain



Ankit Kumar, Saksham Srivastava, Tiansheng Yang, Lu Wang,  
and Rajkumar Singh Rathore

**Abstract** The Internet of Things (IoT) is a revolutionary process that enables seamless communication between electronic devices and electronic devices over the Internet, thus allowing us to be productive and efficient in every aspect of our lives. The Internet of Things leverages smart devices and Internet connectivity to deliver innovative solutions to a variety of challenges faced by industries worldwide. In the proposed project, a temperature sensor is designed using DS18B20 sensor and Arduino microcontroller to carefully monitor and control the water temperature. The system represents a significant advance in water quality monitoring, with applications ranging from environmental monitoring to industrial processes. Temperature sensors play an important role in water leaks by detecting temperature changes over time. They are well versed in monitoring these changes and can detect threats early, take immediate action to reduce risks and maintain stringent water standards. Through this initiative, the Internet of Things continues to drive change worldwide, promote sustainability, and improve the overall quality of life. By using IoT technology, we are not only improving resource management, but also paving the way for a more connected and prosperous future.

**Keywords** Temperature sensing · Arduino micro-controller · Water temperature regulation · Internet of Things (IoT)

---

A. Kumar · S. Srivastava  
Kalinga Institute of Industrial Technology, Bhubaneswar, India

T. Yang (✉)  
University of South Wales, Pontypridd, UK  
e-mail: [tiansheng.yang1@southwales.ac.uk](mailto:tiansheng.yang1@southwales.ac.uk)

L. Wang  
Xi'an Jiaotong-Liverpool University, Wuzhong, Suzhou, China

R. S. Rathore  
Cardiff School of Technologies, Cardiff Metropolitan University, Llandaff Campus, Cardiff, UK  
e-mail: [rsrathore@cardiffmet.ac.uk](mailto:rsrathore@cardiffmet.ac.uk)

## 1 Introduction

Temperature sensors play a pivotal role in water quality monitoring, particularly in the context of ensuring safe and fresh drinking water. With growing concerns about water quality, the need for precise monitoring mechanisms has become increasingly critical. Temperature sensors serve as indispensable tools in this regard, as they enable the detection of changes in temperature over time, providing vital insights into the quality of water. By monitoring temperature variations, water quality professionals can assess the suitability of water for consumption and promptly address any issues that may compromise its safety.

In water quality monitoring, temperature sensors are instrumental in detecting various phenomena that can affect water quality. From identifying thermal pollution, which results from the addition of heat to water bodies due to industrial processes or other sources, to ensuring compliance with regulatory standards regarding water temperature, these sensors play a multifaceted role. Moreover, temperature sensors aid in monitoring water mixing, a crucial aspect in reservoirs, lakes, and other water systems. By tracking temperature changes, it becomes possible to pinpoint areas where remediation efforts are necessary, thereby safeguarding water quality and protecting the environment and public health.

Temperature sensors are indispensable tools in water quality monitoring, offering invaluable insights into changes in water temperature over time. Their ability to detect fluctuations in temperature enables the prevention of thermal pollution, ensures regulatory compliance, facilitates efficient water treatment, and aids in monitoring water mixing. As the focus on water quality intensifies, the importance of temperature sensors in safeguarding the quality of drinking water cannot be overstated. By leveraging temperature sensors, it becomes possible to uphold the integrity and safety of the water we consume, ensuring that it remains fresh and fit for consumption.

Beyond their role in water quality monitoring, temperature sensors contribute significantly to enhancing the efficiency and effectiveness of water treatment processes. By accurately measuring the temperature of water, treatment facilities can optimize the dosage of chemicals and energy required for treatment, thereby reducing costs and minimizing environmental impact. Additionally, temperature sensors play a crucial role in ensuring the proper functioning of equipment within water treatment plants. They help monitor the temperature of various components, such as pumps, valves, and pipelines, detecting any anomalies that may indicate potential equipment failures or inefficiencies. By providing real-time temperature data, these sensors enable operators to take proactive measures to prevent disruptions and maintain the reliability of water treatment operations.

Furthermore, temperature sensors are essential for assessing the impact of climate change on water bodies and ecosystems. As global temperatures rise, water temperatures also increase, affecting aquatic habitats and ecosystems. Temperature sensors play a vital role in monitoring these changes, enabling scientists and policymakers to understand the implications of climate change on water resources and develop strategies to mitigate its effects. By tracking temperature trends over time, researchers can

identify areas that are particularly vulnerable to temperature changes and prioritize conservation efforts accordingly. Additionally, temperature sensors are used in conjunction with other environmental sensors to monitor parameters such as pH, dissolved oxygen, and conductivity, providing a comprehensive understanding of water quality and ecosystem health.

In conclusion, temperature sensors are indispensable tools in water quality monitoring, offering valuable insights into changes in water temperature and facilitating the maintenance of safe and fresh drinking water. From detecting thermal pollution to optimizing water treatment processes and assessing the impact of climate change on aquatic ecosystems, these sensors play a multifaceted role in safeguarding water resources and protecting public health and the environment. As technological advancements continue to improve the accuracy and reliability of temperature sensing technologies, their importance in ensuring the sustainability and resilience of water systems will only continue to grow.

The primary highlights of the paper are as follows:

- Component Selection—Initially, we have selected the best required components such as Micro-controller, i.e., Arduino, Sensors—Water temperature sensor (DS18B20), LCD Display, Relay.
- Basic Prototype—Assembled the selected components into a prototype system capable of monitoring water temperature and regulating it as needed. This involves wiring the sensors to the microcontroller, programming the microcontroller to read sensor data and control heating/cooling elements, and testing the system's functionality.
- Data Transmission Setup—Connected the LCD display to the microcontroller (e.g., Arduino) using appropriate wiring and compatible interface pins. By integrating an LCD display into the IoT device, users can easily monitor temperature data and system status in real-time, enhancing the overall usability and effectiveness of the temperature monitoring and regulation system.
- Result Analysis—After the successful execution of the system, we noted down the result for multiple values in the form of table and graph.

## 2 Literature Survey

The paper [1] addresses the pressing need for advanced technology in aquaculture management, particularly in remote water quality monitoring and computer-controlled intensive culture. The study highlights the significance of real-time monitoring systems in improving aquaculture challenges. Focusing on practices and ensuring food security amidst global temperature sensing as a crucial parameter in water quality monitoring, the research aims to design and develop an efficient mobile-based system capable of providing real-time temperature data. By leveraging RESTful APIs, cloud computing, and IoT technologies, the system enables the collection and dissemination of vital temperature information to aid aquaculture farmers in optimizing fish production cycles and maintaining optimal water conditions. The

methodology involves requirements elicitation, system design, and implementation, utilizing tools such as Raspberry Pi, DS18B20 Digital Temperature sensors, and PostgreSQL database. The study's findings demonstrate the system's effectiveness in accurately and efficiently acquiring real-time water temperature data, thereby supporting decision-making processes and enhancing aquaculture sustainability. The paper [2] presents an innovative approach to temperature control in industrial heating furnaces using a single-chip microcomputer-based system. The system aims to improve temperature control efficiency and reliability by utilizing the advantages of single-chip microcomputers, such as low power consumption, high performance, and ease of production. The design process involves both hardware and software aspects, including sensor selection, signal conversion, A/D conversion, and control signal output. By employing an 80C51 single-chip microcomputer as the core controller, the system achieves constant temperature control through real-time temperature monitoring, comparison with the set temperature range, and automatic adjustment. The paper outlines the hardware and software design details, including unit circuit design, single-chip microcomputer basic system configuration, and software flowchart for signal processing and control. The system's technical specifications include a temperature setting range of 40–90 °C, with a minimum division of 1 °C and a static error of less than or equal to 1 °C. LED digital tube displays are used for temperature visualization, with LED indicators for system status indication. Overall, the intelligent water temperature monitoring system offers a cost- effective and efficient solution for precise temperature control in industrial processes, contributing to enhanced reliability and operational efficiency. The paper [3] addresses the pressing need for effective water quality monitoring systems, particularly in regions like Sub-Saharan Africa where access to clean water is limited. Their study introduces a novel domestic water temperature, pH, and turbidity monitoring system that surpasses traditional manual methods by continuously logging data and providing real-time alerts for potential water quality issues. Leveraging sensors, an Arduino microcontroller, and a NodeMCU board, the system enables seamless data transmission to a cloud platform for analysis and visualization. The results underscore the system's efficacy in tracking water quality parameters, suggesting its potential to enhance water management practices and mitigate waterborne diseases. This work aligns with the broader goal of achieving universal access to safe drinking water, as outlined in the United Nations' Sustainable Development Goals. The paper [4] introduces a novel water temperature measurement system utilizing plastic optical fibers, designed to address the risk of Legionella bacteria growth in water systems. Legionnaire's disease, caused by inhaling water droplets containing harmful bacteria, poses significant health risks, particularly in environments like cooling towers and hot-water systems where the bacteria thrive within specific temperature ranges. Existing monitoring systems face challenges such as high costs, contamination risks, and susceptibility to electromagnetic interference (EMI). The proposed system overcomes these limitations by leveraging the unique properties of plastic optical fibers, which are unaffected by EMI, safe for water use, and cost-effective. The system operates based on macro bending of optical fibers, exhibiting a linear response to temperature changes, enabling accurate and continuous temperature monitoring. Experimental results demonstrate the

system's effectiveness in detecting temperature variations within the critical range for Legionella bacteria growth, highlighting its potential for enhancing water safety in various applications. The paper [5] presents a comprehensive study on the design, calibration, and application of a smart temperature sensor, TS-V1, tailored for high-density water temperature monitoring in estuarine and coastal areas. By addressing the pressing need for precise and cost-effective temperature monitoring solutions, the study fills a crucial gap in the existing literature. The research showcases the development and validation of the TS-V1 sensor against established commercial sensors, providing valuable insights into its accuracy, sensitivity, and stability across various temperature scenarios. Through controlled experiments and in situ deployments in the Pearl River Estuary, the study demonstrates the TS-V1 sensor's capability to capture spatial-temporal variations in surface water and air temperatures with high precision. Additionally, the paper contributes to the field by employing advanced statistical techniques, such as the Generalized Extreme Value (GEV) distribution, to analyze temperature data, enhancing the understanding of thermal dynamics in estuarine ecosystems. Overall, the study significantly advances the field of environmental monitoring by introducing a novel sensor design that offers both reliability and affordability, making it particularly valuable for long-term and high-density temperature monitoring applications in coastal environments. The paper [6] presents water monitoring in fish ponds with respect to temperature involves the regular observation and recording of water temperatures within the pond environment to ensure optimal conditions for fish health and growth. This process typically includes the use of temperature sensors placed at various depths throughout the pond to capture any variations in temperature gradients. Monitoring temperature fluctuations is crucial as it directly affects the metabolic rate, feeding behavior, and overall well-being of the aquatic species. By closely monitoring water temperatures, fish farmers can make informed decisions regarding feeding schedules, stocking densities, and other management practices to maintain an ideal aquatic environment for their fish. Additionally, continuous temperature monitoring can help identify any potential issues such as thermal stratification, oxygen depletion, or the onset of disease, allowing for timely intervention to mitigate risks and ensure the sustainability of the fish farming operation. The paper [7] is a comprehensive resource delving into the pivotal role of water temperature in aquaculture systems. It meticulously examines the multi-faceted effects of temperature on aquatic organisms, encompassing various species like fish, crustaceans, and mollusks. The book elucidates how temperature influences fundamental biological processes such as growth, reproduction, metabolism, and immune function. The paper [8] Water temperature monitoring and regulation IoT devices are invaluable in manufacturing industries for a variety of applications. They can be utilized in processes such as cooling systems for machinery, ensuring optimal temperatures to prevent overheating and maintain equipment performance. Additionally, in chemical manufacturing, these devices help regulate water temperatures for precise control over reactions and product quality. Furthermore, in food and beverage production, maintaining specific water temperatures is crucial for sanitation and product consistency. In all these contexts, IoT devices offer real-time monitoring and control capabilities, enhancing operational efficiency, product quality, and

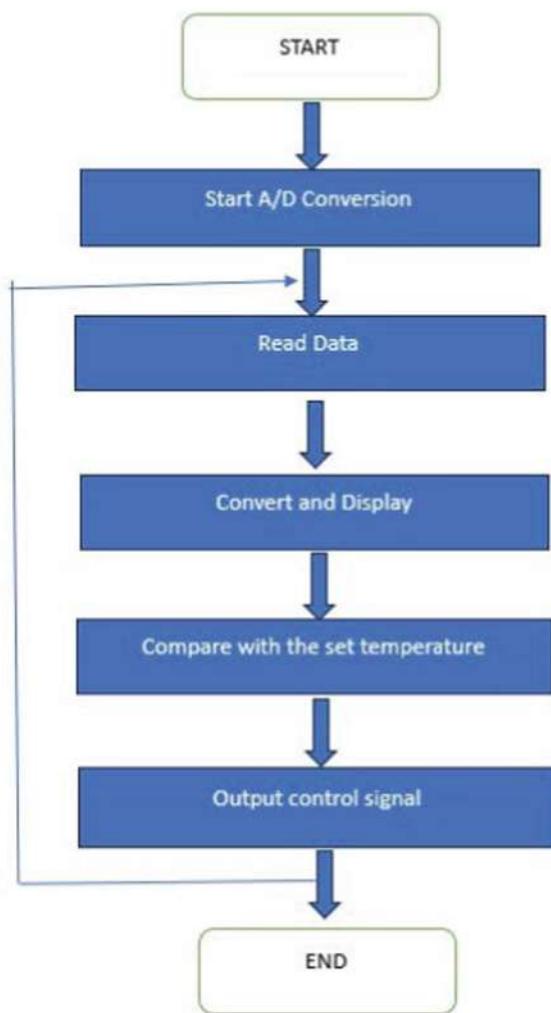
workplace safety in manufacturing industries. The paper [9] likely discusses how IoT sensors are deployed to collect real-time data on various environmental parameters such as air quality, temperature, humidity, and pollution levels. These sensors are likely connected to the Arduino Computer, which processes the data and may perform analysis or trigger actions based on predefined thresholds or conditions. The authors probably explore the advantages of using IoT technology in environmental monitoring, such as improved data accuracy, scalability, and cost-effectiveness. They may also discuss potential applications of such a system in areas like smart cities, agriculture, and industrial monitoring. Overall, this work likely presents a novel approach to environmental monitoring by integrating IoT technology with the Arduino Computer platform, aiming to provide actionable insights for better environmental management and decision-making.

### 3 Proposed Methodology

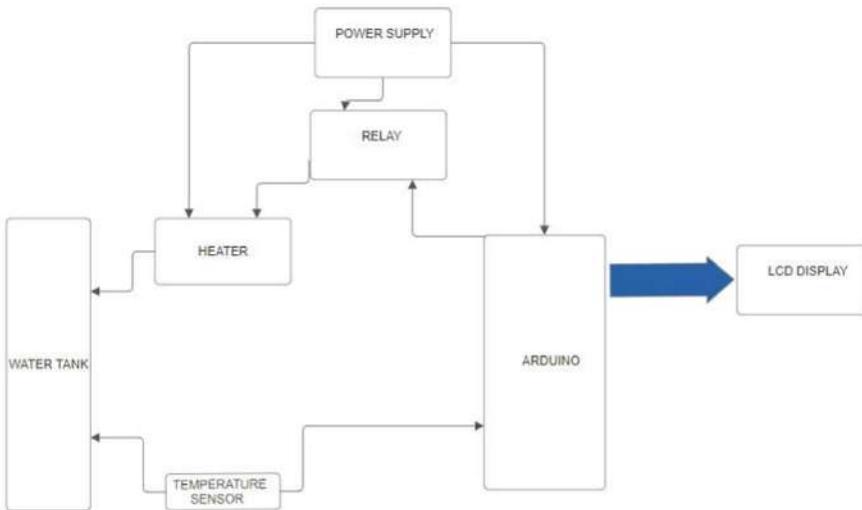
The implementation of the automated water regulation system in energy constraint involves the following methodology. The Arduino Uno microcontroller is connected with the DS18B20 temperature sensor, LCD display (LiquidCrystal\_I2C), and relay for water heater control. Necessary Arduino libraries (OneWire, DallasTemperature, Wire) for sensor communication and LCD display control are installed. The temperature sensor and LCD display in the setup() function of the Arduino sketch are initialized. The pin modes and initial states for controlling the relay (relayPin) are enabled. The temperature data from the DS18B20 sensor using the OneWire and DallasTemperature libraries are aggregated [10–12]. The temperature readings on the LCD display in both Celsius and Fahrenheit units are displayed. The control logic to activate or deactivate the relay based on predefined temperature thresholds (threshold\_2) for water heating is implemented. The error handling mechanisms to manage sensor disconnections or communication failures is applied and the error messages on the LCD display and take appropriate actions to ensure system reliability is displayed [13–15]. The iterative testing is conducted to verify temperature monitoring accuracy and relay control functionality under different temperature conditions. The system behavior is validated against expected outcomes and adjust control parameters as needed. Finally, all components into a functional system setup are integrated ensuring proper wiring and connectivity.

Figure 1 highlights the work flow model diagram. The process begins by initiating the Analog-to-Digital (A/D) conversion process. This converts the analog signal from the temperature sensor (DS18B20) into a digital value that the microcontroller can understand. The Arduino reads the digital temperature value provided by the temperature sensor. The digital temperature value is converted into a readable temperature unit (likely Celsius) and then displayed on the LCD screen. The Arduino compares the measured temperature with the preset threshold temperature (likely 50 °C). This refers to turning the heater on or off depending on the temperature comparison. The process loops back to the beginning to continuously monitor the water temperature.

**Fig. 1** Process flow diagram of the model



The system architecture of the model as shown in Fig. 2 encompasses the interconnection and interaction of its various hardware and software components. The block diagram illustrates the flow of data and signals within the system. At the core of the architecture is the Arduino microcontroller, which serves as the central processing unit. Connected to the microcontroller is the DS18B20 temperature sensor, responsible for measuring water temperature. The sensor communicates temperature data to the microcontroller, which processes the information and sends it to the LCD display for visualization. Additionally, the microcontroller interfaces with a relay, enabling it to control the heater based on the temperature readings. The entire system is assembled on a breadboard, facilitating the physical connection of components and ensuring their proper functioning. This systematic arrangement ensures efficient



**Fig. 2** System architecture of the model

temperature monitoring, display, and control, contributing to the project's overall effectiveness in detecting and mitigating thermal pollution in industrial waste water.

#### 4 Implementation Analysis

The development of an IoT-based Automated Water Monitoring and Regulation System is addressed in this study. In this research, an Arduino is utilized as a controller, with an ultrasonic sensor to sense the water level of the tank, and then a relay switch is enabled or disabled on basis of a certain water level, allowing the water pump to water the tank. A GSM unit is used to track the water level.

Table 1 illustrates the power consumption analysis of the proposed model. As seen, water module intakes the maximum voltage of 220 V and 3.24 A current. The power is computed by multiplying the voltage level with its current. The heater generates the highest power of 1440 W.

**Table 1** Power consumption of equipment

Device	Voltage (V)	Current (A)	Power (W)
Water module	220 AC	3.24	712.8
Arduino	4.9 AC	0.35	1.715
Relay unit	4.9 DC	0.12	0.588
LCD	5.0 DC	0.14	0.7
Heater	120 DC	12	1440

**Fig. 3** Reliability degree analysis of the proposed model

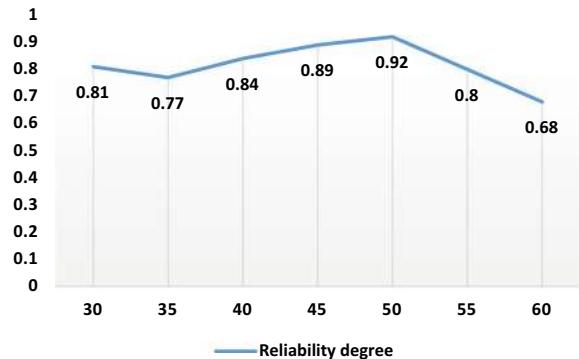


Figure 3 displays the degree of reliability of the discussed automated model. It is observed that when the temperature is preset at 50° then the reliability recorded of 0.92 is maximum while it is least at 0.68 for specified temperature of 60°.

## 5 Conclusion

The development of an IoT-based Automated Water Monitoring and Regulation System marks a significant leap forward in water management technology. Traditional water heating setups often falter due to their inability to monitor and regulate in real-time, leading to substantial energy wastage and escalated operational expenses across residential, commercial, and industrial domains. This endeavor tackles these challenges head-on by deploying a smart system endowed with the capability to dynamically control water temperature based on precise sensor data. At its core, this project leverages the integration of an Arduino-Uno microcontroller, DS18B20 temperature sensor, LCD screen, and water heater, showcasing a promising avenue to optimize energy consumption, curtail operational expenditures, and mitigate environmental footprints. By orchestrating the interplay between these components, our system adeptly fine-tunes water temperature in accordance with pre-established thresholds, ensuring judicious utilization of heating resources. Such precision aligns seamlessly with sustainable principles governing water heating applications, fostering a more ecologically responsible approach to resource utilization. Beyond mere operational enhancements, the implications of this innovation reverberate across diverse sectors. In residential settings, homeowners stand to benefit from reduced utility bills and heightened convenience, while commercial establishments and industrial facilities can streamline their operations, bolstering productivity and profitability. Moreover, the ripple effects extend to environmental conservation efforts, as the judicious use of energy resources contributes to mitigating carbon footprints and promoting ecological stewardship.

In essence, the IoT-based Automated Water Monitoring and Regulation System heralds a new era of efficiency and sustainability in water management, epitomizing the transformative potential of technology in safeguarding our precious natural resources for generations to come.

## References

1. Bokingkito PB, Llantos OE (2017) Design and implementation of real-time mobile-based water temperature monitoring system. *Procedia Comput Sci* 124:698–705
2. Panda A, Mishra S, Rathore R, Obaid AJ, Alkhafaji MA (2024) Transmission through networks using federated learning approach. In: Proceedings of fifth doctoral symposium on computational intelligence: DoSCI 2024, vol 4. Springer Nature, p 121
3. Nanyanzi DR, Ocen GG, Omara T et al (2021) Design and assembly of a domestic water temperature, pH and turbidity monitoring system. *BMC Res Notes* 14:161. <https://doi.org/10.1186/s13104-021-05578-9>
4. Wang B, Cai H, Jia Q, Pan H, Li B, Fu L (2023) Smart temperature sensor design and high-density water temperature monitoring in Estuarine and coastal areas. *Sensors* 23:7659. <https://doi.org/10.3390/s23177659>
5. Ngueku BB (2015) Water monitoring in fish ponds. *Int J Fish Aquat Stud* 2:31–32
6. Boyd CE, Tucker CS (1998) Pond aquaculture water quality management. Springer, US
7. Rathore RS, Sangwan S, Mazumdar S, Kaiwartya O, Adhikari K, Kharel R, Song H (2020) W-GUN: whale optimization for energy and delay-centric green underwater networks. *Sensors* 20(5):1377
8. Ravindran MA, Nallathambi K, Vishnuram P, Rathore RS, Bajaj M, Rida I, Alkhayyat A (2023) A novel technological review on fast charging infrastructure for electrical vehicles: challenges, solutions, and future research directions. *Alex Eng J* 82:260–290
9. Kumar M, Kumar S, Kashyap PK, Aggarwal G, Rathore RS, Kaiwartya O, Lloret J (2022) Green communication in internet of things: a hybrid bio-inspired intelligent approach. *Sensors* 22(10):3910
10. Sahoo S, Mishra S, Brahma B, Barsocchi P, Bhoi AK (2024) SSO-CCNN: a correlation-based optimized deep CNN for brain tumor classification using sampled PGGAN. *Int J Comput Intell Syst* 17(1):1–18
11. Mishra S, Chaudhury P, Tripathy HK, Sahoo KS, Jhanjhi NZ, Hassan Elnour AA, Abdelmaboud A (2024) Enhancing health care through medical cognitive virtual agents. *Digital Health* 10:20552076241256732
12. Gairola AK, Kumar V, Singh GD, Bajaj M, Rathore RS, Mahmoud MH, El-Shafai W (2024) Efficient deep learning fusion-based approach for brain tumor diagnosis. *Traitement du Signal* 41(5)
13. Das S, Sarkar B, Mishra S, Rathore RS, Faris NN (2024) A real-time sensory model for consciousness tracking with alert notification. In: Doctoral symposium on computational intelligence. Springer Nature Singapore, Singapore, pp 295–302
14. Mishra S, Chakraborty S, Sahoo KS, Bilal M (2023) Cogni-Sec: A secure cognitive enabled distributed reinforcement learning model for medical cyber–physical system. *Internet Things* 24:100978
15. Mishra S, Volety DR, Bohra N, Alfarhood S, Safran M (2023) A smart and sustainable framework for millet crop monitoring equipped with disease detection using enhanced predictive intelligence. *Alex Eng J* 83:298–306

# Deep Neural Network with Stochastic LWTA (DNN-S-LWTA) for Adversarial Machine Learning



Ms. Soumya and Robin Rohit Vincent

**Abstract** Machine Learning (ML) with attackers has been discussed in the recent domain of adversarial ML. It consists of various attacks that try to weaken the accuracy of ML models. Adversarial examples are datasets which contain suspiciously generated deviations to deceive an existing model leading to incorrect classification results. Any little modification in the original dataset may damage the ML model. To provide a better solution for this problem, a Deep Neural Network (DNN) model with Stochastic Local-Winner-Takes-All (S-LWTA) activations, is proposed. In this work, adversarial samples are generated for training, by integrating the original dataset with attack samples. Experimental results exhibit that the proposed DNN-S-LWTA model achieves better accuracy, when compared to existing ML models.

**Keywords** Adversarial Machine Learning · Deep Neural Networks (DNNs) · Local-Winner-Takes-All (LWTA) · Accuracy

## 1 Introduction

Internet of Things (IoT) is commonly considered as a network consisting interconnected objects deployed with sensors for collecting the data, communicating, and executing complex tasks. The applications of IoT range in variety of domains such as smart homes, smart traffic, vehicular network, environment monitoring, and border security [1].

Most of the IoT applications enforce strict demands corresponding to security and reliability. Hence assuring secure and reliable communication of accurate data within an IoT network is the chief requirement [2]. Information security within an

---

Ms. Soumya (✉)

Department of Computer Science Engineering, School of Computer Science Engineering and Information Science, Presidency University, Bengaluru, Karnataka, India  
e-mail: [ms.soumyaphd@yahoo.com](mailto:ms.soumyaphd@yahoo.com)

R. R. Vincent

Computer Science Engineering, Presidency University, Bengaluru, Karnataka, India

IoT network face various issues, such as the secure transmission of data and the secure functioning of data centers [3].

To handle the large continuous volumes of data generated from IoT devices, ML algorithms provide automatic solutions for data processing, analyzing and decision-making. In spite of its benefits, ML has many loopholes that might be utilized by dishonest users. ML containing attackers is analyzed by the latest developed model of Adversarial Machine Learning (AML) [4].

Deep Learning (DL) model has become familiar in the fields of object detection, pattern matching and analysis. Though it has succeeded, DNN models have the risk of adversarial attacks. Adversarial examples are datasets which contain suspiciously generated deviations to deceive an existing model leading to incorrect classification results. Any little modification in the original dataset may damage the ML model. These adversarial models can be created by applying *Adversarial Training*, where a ML or DL model is trained with both original and adversarial data [5, 6].

The major objectives of this research work include:

- (i) To perform adversarial training using customized DNN model to predict the attacks accurately.
- (ii) To evaluate the performance of S-LWTA with other classifiers models.

To meet these objectives this paper design DNN with stochastic LWTA activations (DNN-S-LWTA) model.

## 2 Related Works

Luo et al. [7] introduced an ML technique that focuses on partial-model attacks within the data aggregation stage of IoT. This attack only requires control over lesser fraction of the IoT devices. Their research outcomes clearly illustrate the potential of this adversary to affect the decision-making process in data gathering even with minimal control over the IoT devices.

Yang et al. [8] presented a robust training model for classifying the IoT data based on encoder and decoder. The encoding is applied over the original training dataset, which are then used to rebuild the original time series data through this model. The ResNet model trained on the reconstructed dataset exhibits increased resilience against adversarial attacks.

Qiu et al. [9] devised a novel attack strategy targeting DL-based Network Intrusion Detection Systems (NIDS) in the IoT environment. They use model extraction to copy the black-box model with a minimal set of training samples. A saliency map is then employed to reveal the effect of each packet parameter on the detection outcomes, allowing for the efficient generation of adversarial examples using conventional techniques.

Vitorino et al. [10] introduced a methodology for assessing adversarial robustness, particularly focusing on realistic adversarial evasion attacks. Conditional adversarial

samples were generated using the adaptive perturbation pattern method (A2PM), and evasion attacks were executed against the models of regular and adversarial training.

While different attack strategies have been extensively examined, there is a notable shortage of research focusing on defense mechanisms to enhance the resilience of ML-based classifiers. Furthermore, there is a lack of customized mechanisms aimed at safeguarding IoT from AML attacks, particularly in applications like object detection.

### 3 Proposed Methodology

Recently, the DL field has turned its attraction LWTA model. Under this model, a single neuron can be active at a given time in a block, while the remaining ones are deactivated [11].

This work proposes a new DL model which is customized to address adversarial deep learning. It employs DNN with S-LWTA activations. The cyber-attacks used in the training dataset are Brute-force attack, Malformed attack, and SlowITe attack.

The dataset is generated from the raw packet traces which are obtained by conducting real-time simulations using Contiki/Cooja simulator. The legitimate dataset is obtained by conducting experiments without any network intrusion attacks. Then an attack dataset is obtained by introducing malicious nodes in the network, causing Brute-force attack, Malformed attack, and SlowITe attack. Then the resultant adversarial dataset is obtained by mixing the legitimate dataset and the attack dataset in the ratio of 60:40.

#### 3.1 S-LWTA Network

Consider an input, denoted as  $x \in R^J$ , fed into a traditional DNN layer having  $N$  hidden neurons with a weight matrix  $W \in R^{J \times N}$ . Each hidden neuron, labeled as  $k$ , in the layer, conducts an inner product computation  $h_n$ ,

$$h_n = w_n^T x = \sum_{i=1}^I w_{in} \cdot x_i \in R \quad (1)$$

Consequently, the output vector of the last layer, denoted as  $y \in R^N$ , is created by concatenating the non-linear activations of every individual hidden unit, creating

$$y = [y_1, \dots, y_N], \text{ where } y_n = \sigma(h_n)$$

In a fully connected layer with LWTA, individual non-linear units are substituted by  $U$  linear participating units, assembled combinedly, in a LWTA block. The associated weights are structured as 3D matrix,  $W \in R^{I \times A \times U}$ , where  $A$  is the count of LWTA blocks in a layer.  $x$  is offered to each block and unit within them. In this framework, the  $u$ th competing unit within the  $a$ th block calculates its activation  $h_{a,u}$  through a standard inner product computation:

$$h_{a,u} = w_{a,u}^T x = \sum_{i=1}^I w_{j,a,u} \cdot x_i \in R \quad (2)$$

Of the  $U$  units in an LWTA block, one unit is emerged as the winner, which carries its linear activation to the subsequent layer, whereas the remaining units yield 0 values. Consequently, the output of an LWTA layer,  $y \in R^{A,U}$ , comprises  $A$  subvectors  $y_a \in R^U$ , each related to an LWTA block and featuring a solitary non-zero entry. This competitive process integrally produces a sparse representation, as all units, barring one in each block, yield zero outputs.

To signify the victorious unit within each blocks, a S-LWTA layer is formed. So, the output  $y$  of the  $(a, u)$ th component  $y_{a,u}$  of a stochastic LWTA layer is given by:

$$y_{a,u} = \varepsilon_{a,u} \sum_{i=1}^I w_{j,a,u} \cdot x_i \in R \quad (3)$$

where  $\varepsilon_{a,u}$  is the  $\varepsilon_a$ 's  $u$ th component that holds the  $a$ th subvector of  $\varepsilon, t$ . The likelihood of unit for being the winner in a specific block improves with its higher linear response in that block. This leads to:

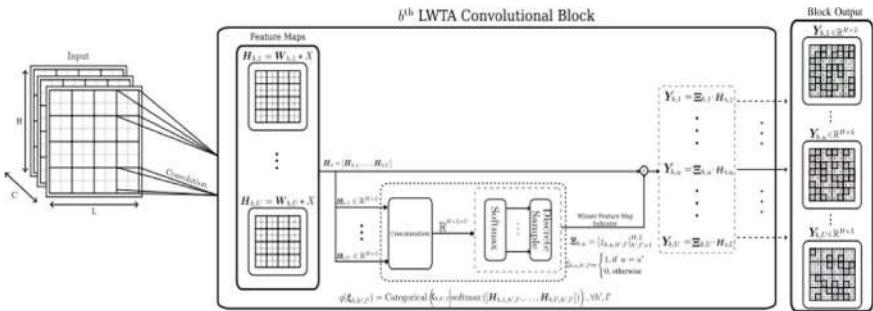
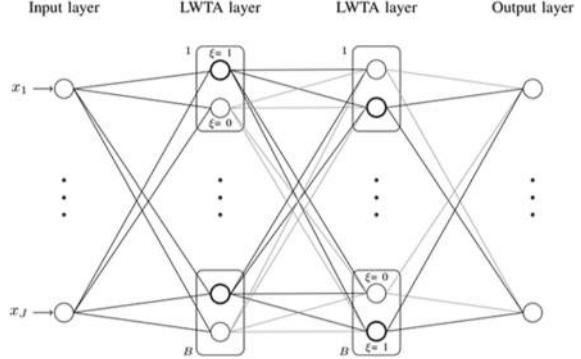
$$q(\varepsilon_a) = \text{Categorical}\left(\varepsilon_a | \text{softmax}\left(\sum_{i=1}^I [w_{i,a,u}]_{u=1}^U \cdot x_i\right)\right) \quad (4)$$

where  $[w_{i,a,u}]_{u=1}^U$  is the vector joining of the set  $\{w_{i,a,u}\}_{u=1}^U$ .

Each S-LWTA layer contains LWTA blocks, as depicted in Fig. 1, where rectangles are the blocks, and circles are the participating units. The winners are denoted by bold contours ( $\varepsilon = 1$ ). Figure 2 shows the proposed stochastic convolutional LWTA block [12].

With an input  $X$ , contest occur among feature maps on a position-specific basis. In a specific position, only the feature map determined as the winner has a non-zero entry.

**Fig. 1** Competition-based modeling approach



**Fig. 2** Stochastic convolutional LWTA block

### 3.2 Training

Considering the data  $D = \{X_j, Y_j\}$  for  $j = 1-M$ , the categorical cross-entropy between the data labels  $Y_j$  and the class probabilities  $f(X_j; \hat{\varepsilon})$  are defined, generated by the penultimate Softmax layer, as  $\text{CE}(Y_j, f(X_j; \hat{\varepsilon}))$ , where  $\hat{\varepsilon}$  represents sample instances of all latent variables  $\varepsilon$  across all layers. The output class probabilities are depending on the stochastic winner selection process in each layer. Consequently, the Evidence Lower Bound (ELBO) is formulated as follows:

$$E = - \sum_{X_j, Y_j \in D} (\text{CE}(Y_j, f(X_j; \hat{\varepsilon})) - \text{KL}(q(\varepsilon) \| p(\varepsilon))) \quad (5)$$

## 4 Experimental Results

### 4.1 Dataset Details

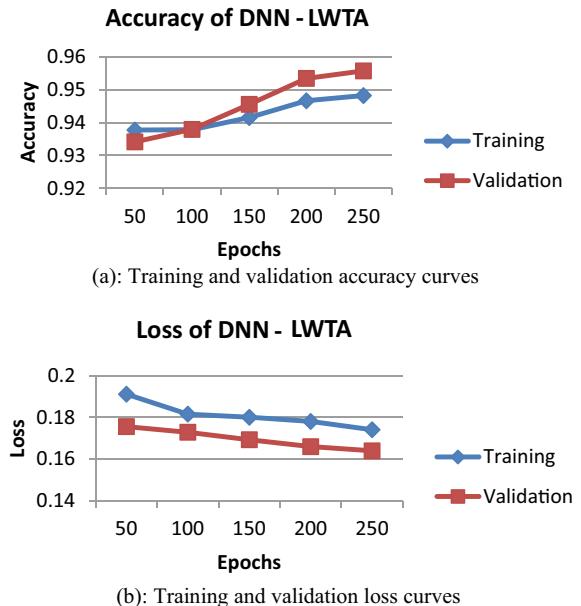
In this work, the MQTT dataset is used, which is based on the Message Queuing Telemetry Protocol (MQTT) protocol of IoT network. The dataset is created by combining the legitimate dataset with cyber-attacks against the network. The dataset was then split into training and testing sets as 70:30%. The cyber-attacks used in the training dataset are Brute-force attack, Malformed attack, and SlowITe attack.

### 4.2 Results

In this section, adversarial training is conducted using the proposed DNN-S-LWTA algorithm. Figure 3 shows the training and validation curves of accuracy and loss obtained by DNN-S-LWTA.

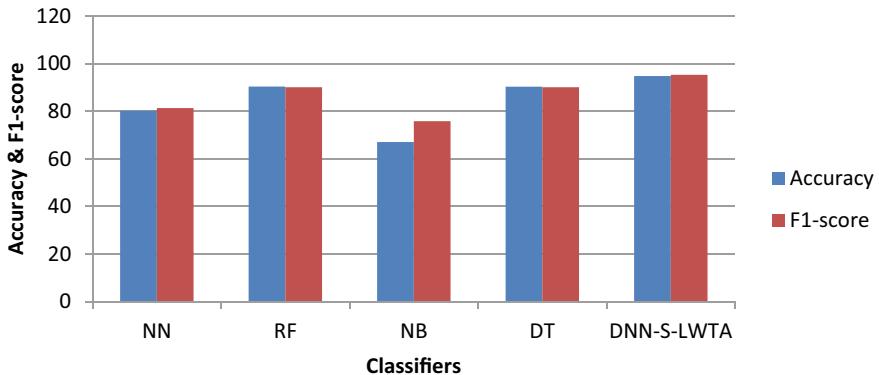
Along with the proposed DNN-LWTA model, other classifiers Neural Network (NN), Random Forest (RF), Naive Bayes (NB), and Decision Tree (DT) are used to predict the attack correctly. Then, the accuracy and F1-score of these models are compared and the outcomes are presented in Table 1 and Fig. 4.

**Fig. 3** **a** Training and validation accuracy curves, **b** training and validation loss curves



**Table 1** Prediction results of DNN-S-LWTA and other classifiers

Metrics	NN	RF	NB	DT	DNN-S-LWTA
Accuracy	80.27	90.38	67.08	90.31	94.82
F1-score	81.35	90.12	75.81	90.08	95.31

**Fig. 4** Prediction results of classifier models

As we can see from Table 1 and Fig. 4, the proposed DNN-S-LWTA model attains highest accuracy around 94.8% and F1-score around 95.3%, when compared to the other classifiers.

## 5 Conclusion

A DNN-S-LWTA model has been proposed in this paper to address adversarial robust DL. An adversarial dataset is created for training, by combining the legitimate dataset with cyber-attacks. The dataset was then split into training and testing sets as 70% and 30%, respectively. The cyber-attacks used in the training dataset are Brute-force attack, Malformed attack, and SlowITe attack. Experimental outcomes have shown that the accuracy of DNN-S-LWTA model is superior to the ML models NN, RF, NB, and DT.

## References

- Bütün I, Osterberg P, Song H (2020) Security of the internet of things: vulnerabilities, attacks, and countermeasures. *IEEE Commun Surv Tutorials* 22(1):616–644
- Sisinni E, Saifullah A, Han S, Jennehag U, Gidlund M (2018) Industrial internet of things: challenges, opportunities, and directions. *IEEE Trans Ind Inf* 14(11):4724–4734

3. Neshenko N, Bou-Harb E, Crichigno J, Kaddoum G, Ghani N (2019) Demystifying IoT security: an exhaustive survey on IoT vulnerabilities and a frst empirical look on internet-scale IoT exploitations. *IEEE Commun Surv Tutorials* 21(3):2702–2733. <https://doi.org/10.1109/COMST.2019.2910750>
4. Srivastava A, Gupta S, Quamara M, Chaudhary P, Aski VJ (2020) Future IoT-enabled threats and vulnerabilities: state of the art, challenges, and future prospects. *Int J Commun Syst* 33:12
5. Martins N, Cruz JM, Cruz T, Henriques Abreu P (2020) Adversarial machine learning applied to intrusion and malware scenarios: a systematic review. *IEEE Access* 8:35403–35419
6. Apruzzese G, Andreolini M, Ferretti L, Marchetti M, Colajanni M (2021) Modeling realistic adversarial attacks against network intrusion detection systems. *Digit Threat Res Prac* 1
7. Luo Z, Zhao S, Lu Z, Sagduyu YE, Xu J (2020) Adversarial machine learning based partial-model attack in IoT. arXiv. <https://doi.org/10.1145/3395352.3402619>
8. Yang Z, Abbasi IA, Algarni F, Ali S, Zhang M (2021) An IoT time series data security model for adversarial attack based on thermometer encoding. *Secur Commun Netw* 2021
9. Qiu H, Tian D, Zhang T, Lu J, Memmi G, Qiu M (2021) Adversarial attacks against network intrusion detection in IoT systems. *IEEE Internet Things J* 8(13):10327–10335. <https://doi.org/10.1109/jiot.2020.3048038>
10. Vitorino J, Praça I, Maia E (2023) Towards adversarial realism and robust learning for IoT intrusion detection and classification. *Ann Telecommun* 78:401–412
11. Panousis KP, Chatzis S, Alexos A, Theodoridis S (2021) Stochastic local winner-takes-all networks enable profound adversarial robustness. In: Bayesian deep learning workshop, NeurIPS 2021
12. Panousis KP, Chatzis S, Alexos A, Theodoridis S (2021) Local competition and stochasticity for adversarial robustness in deep learning. In: Proceedings of the 24th International conference on artificial intelligence and statistics (AISTATS) 2021, vol 130. PMLR, San Diego, California, USA

# Stock Price Prediction Using Arithmetic Optimizer-Assisted LSTM Model



P. V. Bhuvaneshwari, Radhakrishnan Vignesh, and H. B. Asif Mohamed

**Abstract** The stock market functions as a financial marketplace where shares of publicly sorted companies are transacted through buying and selling activities. Serving as an economic indicator, it reflects the performance of businesses and the entire economy. Stock prices are affected by the laws of supplies and demands. This paper presents a deep learning (DL)-based long short-term memory (LSTM) model in time series dataset for effective stock price prediction. A metaheuristic arithmetic optimization algorithm is applied in order to fine tune the hyperparameters of LSTM and to improve the accuracy of stock price prediction. The DJIA dataset is considered in the experiments. Results have shown that the proposed model attains higher  $R^2$  and MSE values around 0.891 and 0.02, respectively.

**Keywords** Long short-term memory · Arithmetic optimization · Stock market price prediction

## 1 Introduction

A stock market is considered as a place, where users can purchase and exchange company stocks, aiming at financial gains. It is important for the economic welfare of the country by providing both company benefits and larger business scope [1]. The stock market operates at various exchanges which assist in trading of stocks, which provide insights of performance of the company and business environment [2].

---

P. V. Bhuvaneshwari (✉) · R. Vignesh · H. B. Asif Mohamed  
Department of CSE, Presidency University, Itgalpur, Rajanakunte, Yelahanka, Bengaluru,  
Karnataka, India  
e-mail: [bhuvaneshwaripvphd@gmail.com](mailto:bhuvaneshwaripvphd@gmail.com)

R. Vignesh  
e-mail: [Vignesh.r@presidencyuniversity.in](mailto:Vignesh.r@presidencyuniversity.in)

H. B. Asif Mohamed  
e-mail: [asif.mohamed@presidencyuniversity.in](mailto:asif.mohamed@presidencyuniversity.in)

Different types of stocks such as Warrants Stock (WS), Common Stock (CS), and Preferred Stock (PS) exist in the market [3]. Recently, DL model has attracted significant importance in the financial domain, especially in the field of stock markets. Predicting the stock market prices has usually been a crucial challenge because of the market's inconsistency and complexity. With the emergent of advanced mechanisms and the availability of large volumes of financial datasets, DL has become an efficient tool to handle this challenge [4].

DL consists of neural networks with various layers, which has the ability to identify the complex patterns and associations among the data. Hence DL techniques are well suited for stock price prediction problems, where patterns can be found in the collected data which involve trading volumes, sentiment analysis, and other financial predictions. LSTM is a special type of DL model designed for analysis and knowledge extraction of data with long-term dependencies. LSTM assists in determining the patterns and trends within the time series stock market data [5].

## 2 Literature Survey

Ingle et al. [6] have presented ensemble DL model for stock price prediction and Term Frequency-Inverse Document Frequency (TF-IDF) technique for feature extraction. This TF-IDF was utilized to count word score, and the accuracy achieved was 85%

Gülmез and Burak [7] presented a LSTM with artificial rabbit's algorithm (ARA) for SPP. For the performance enhancement, the network LSTM was combined with ARA. The dataset DJIA was utilized and the results were generated by varying the value of the ticker and attained higher MSE and MAE values.

Liu et al. [8] presented a LSTM with particle swarm optimizer (PSO) for SPP. Initially, the empirical wavelet transformation (EWT) was utilized for preprocessing, and the outlier robust extreme learning machine (ORELM) was utilized for post-processing stage. At last, the LSTM with PSO was utilized for enhancing accuracy.

Gunduz and Hakan [9] presented Variational autoEncoder (VAE) for SPP. Here, the features were selected by the recursive feature elimination, and the classification process was performed. Finally, accuracy achieved was 0.675 on the eight banking stocks.

Rezaei et al. [10] developed SPP using DL model and decomposition of frequency. Initially, the empirical mode decomposition (EMD) was used for extracting features. Then, the DL model CNN with LSTM was used for SPP. Finally, MAE and RMSE values achieved were 138 and 185.6 on Dow jones dataset.

Patil et al. [11] forecasted SPP and utilized two classifications like Rider Deep LSTM and crow search Deep RNN. It was developed by incorporating the Rider optimization along with Deep LSTM and the crow search optimization with Deep RNN. Finally, the MSE and RMSE values achieved were 0.018 and 0.132.

### 3 Proposed Methodology

Figure 1 presents the workflow of the proposed SPP which includes pre-processing, automatic feature extraction, and SPP using the LSTM with AO. The proposed method combines LSTM with AO to enhance the accuracy of SPP. This hybrid approach leverages the strengths of LSTM networks in capturing temporal dependencies and the efficiency of AO in tuning model parameters for improved performance.

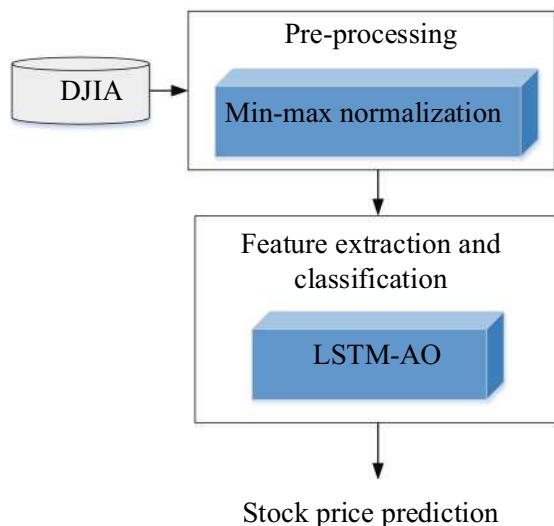
#### 3.1 Preprocessing

Initially, the dataset is preprocessed by the min-max normalization and this process adjusts the values in the dataset to a common scale, typically between 0 and 1.

$$g = \frac{f - f_{\min}}{f_{\max} - f_{\min}}, \quad (1)$$

where  $g$  is the normalized value,  $f$  is the value to be determined, and  $f_{\min}$  and  $f_{\max}$  are the minimum and maximum values.

**Fig. 1** Workflow of the proposed SPP



### 3.2 Feature Extraction and SPP

The preprocessed time series data into the LSTM network and it capture the short-term dependencies in the stock price data. Generate the final stock price prediction. Find the objective function as the Mean Squared Error (MSE) between the predicted and actual stock prices.

$$\text{objective} = \text{Minimize}(\text{MSE}). \quad (2)$$

LSTM belonging to a distinctive category of RNN is employed for extracting and predicting of stock market. The LSTM model incorporates three primary gates: the input gate  $i_l$ , forget gate  $f_l$ , and output gate  $o_l$  as shown in Fig. 2. The aim of the  $f_l$  is to get  $C_{l-1}$  the result from prior unit and guarantee which segment of  $C_{l-1}$  is for retaining and forgetting.

$$f_l = \sigma(W_f[h_{l-1}x_l] + b_f). \quad (3)$$

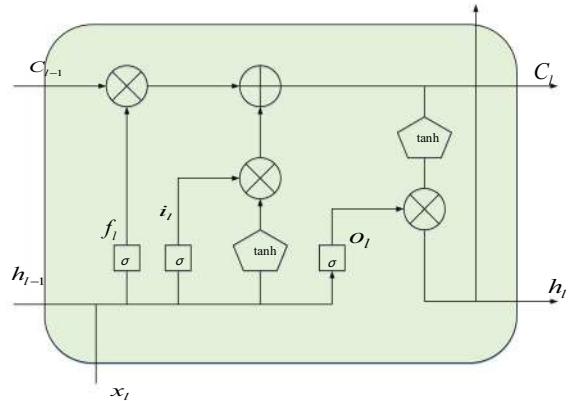
The memory gate  $C_l$  has two layers like sigmoid and tanh, and they are expressed as:

$$i_l = \sigma(W_i[h_{l-1}x_l] + b_i), \quad (4)$$

$$\tilde{C}_l = \tanh(W_C[h_{l-1}x_l] + b_C), \quad (5)$$

$$C_l = f_l \times C_{l-1} + i_l \times \tilde{C}_l. \quad (6)$$

**Fig. 2** LSTM model



The  $o_l$  plays a role in determining the cell state result segment, and the cell state undergoes processing by the tanh. The resulting values are multiplied to obtain the final information intended for output.

$$o_l = \sigma(W_o[h_{l-1}x_l] + b_o), \quad (7)$$

where  $W_i$ ,  $W_f$ ,  $W_o$  and  $b_i$ ,  $b_f$ ,  $b_o$  are the weighting matrices and bias values;  $h_{l-1}$  is the prior output of cell,  $x_l$  is the present cell. Then, to optimize the hyperparameters of LSTM, the algorithm arithmetic optimizer (AO) is utilized. Initialize the parameters of the LSTM network (weights and biases) randomly. The basic arithmetic operations are utilized for exploration and exploitation stages. In AO, the populations are initially presented by the following expression:

$$y = ll + (ul - ll) \times r, \quad (8)$$

where  $ul$ ,  $ll$  and  $r$  are the upper, lower limits and random number. The decision between exploration and exploitation can be made which is based on the outcome of the MOA (Math-Optimized Acceleration). MOA can be calculated as:

$$\text{MOA}(C\_it) = \min + C\_it \times \left( \frac{\max - \min}{M\_it} \right), \quad (9)$$

where  $\text{MOA}(C\_it)$  is the current iteration, and  $\max$  and  $\min$  are the maximum and minimum iterations. The operators like  $D$  and  $M$  are used for performing exploration, and it is given as:

$$y_{j,k}(t+1) = \begin{cases} b(y_k) \div (\text{MOP} + \alpha) \times ((ul_k - ll_k) \times \beta + ll_k) & \text{rand} < 0.5 \\ b(y_k) \times (\text{MOP}) \times ((ul_k - ll_k) \times \beta + ll_k) & \text{elsewhere} \end{cases}, \quad (10)$$

where  $y_{j,k}(t+1)$  is the  $(t+1)$  iteration,  $b(y_k)$  is the fitness, and  $\beta$  is the whole number. The term MOP is given as:

$$\text{MOP} = 1 - \frac{t^{1/\theta}}{M\_it^{1/\theta}}, \quad (11)$$

where  $\theta$  is the essential variable. The operators like  $A$  and  $S$  are used for performing exploitation in which the  $\alpha$  is the constant value and it is given as:

$$y_{j,k}(t+1) = \begin{cases} b(y_k) - (\text{MOP}) \times ((ul_k - ll_k) \times \alpha + ll_k) & \text{rand } 1 < 0.5 \\ b(y_k) + (\text{MOP}) \times ((ul_k - ll_k) \times \alpha + ll_k) & \text{elsewhere} \end{cases}. \quad (12)$$

Algorithm 1 shows the pseudocode of the LSTM with AO.

**Algorithm 1:** Pseudocode of the LSTM with AO**Input:** weights, biases and arithmetic operators**Output:** Optimal value**While**  $t \leq \text{Max\_iter}$ 

Compute the objective by the Eq. (2)

MOA is updated by the Eq. (9)

MOP is updated by the Eq. (11)

**if**  $\text{rand} < 0.5$  **then**        Update the exploration by the  $b(y_k) \div (MOP + \alpha) \times ((ul_k - ll_k) \times \beta + ll_k)$     **else**        Update the exploration by the  $b(y_k) \times (MOP) \times ((ul_k - ll_k) \times \beta + ll_k)$     **if**  $\text{rand } 1 < 0.5$  **then**        Update the exploitation by the  $b(y_k) - (MOP) \times ((ul_k - ll_k) \times \alpha + ll_k)$     **else**        Update the exploitation by the  $b(y_k) + (MOP) \times ((ul_k - ll_k) \times \alpha + ll_k)$     **end if****end if****end for**

Obtain best solution

## 4 Result Analysis

The metrics such as Mean Square Error (MSE), Root MSE (RMSE), Mean Absolute Error (MAE), and  $R^2$  values are computed, which are defined by the following equations:

$$\text{MSE} = \frac{1}{l} \sum_{k=1}^l (y_k - \hat{y}_k)^2, \quad (13)$$

$$\text{RMSE} = \sqrt{\frac{1}{l} \sum_{k=1}^l (y_k - \hat{y}_k)^2}, \quad (14)$$

$$\text{MAE} = \frac{1}{l} \sum_{k=1}^l |y_k - \hat{y}_k|, \quad (15)$$

$$R^2 = 1 - \frac{(y_k - \hat{y}_k)^2}{(y_k - \bar{y}_k)^2}, \quad (16)$$

where  $l$ ,  $y_k$ ,  $\hat{y}_k$ , and  $\bar{y}_k$  are the observations, actual, predicted, and mean actual value.

## 4.1 Dataset

This work considers the DJIA dataset [12], and the time range of the data is about January 4, 2009, to December 31, 2019. It is a stock market index which traces 30 major public limited companies of USA.

## 4.2 Comparative Analysis

Following section states the comparative analysis of the methods like RNN, LSTM, BILSTM, and the proposed LSTM with AO.

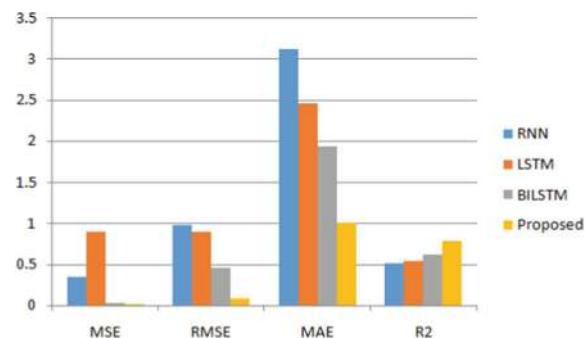
The comparative performance of different models is depicted in Table 1 and Fig. 3. The methods like RNN, LSTM, BILSTM and the proposed LSTM with AO with respect to the measures like MSE, RMSE, MAE and  $R^2$ . Here, the proposed LSTM with AO achieved better MSE of 0.01, RMSE of 0.08, MAE of 1, and  $R^2$  of 0.783.

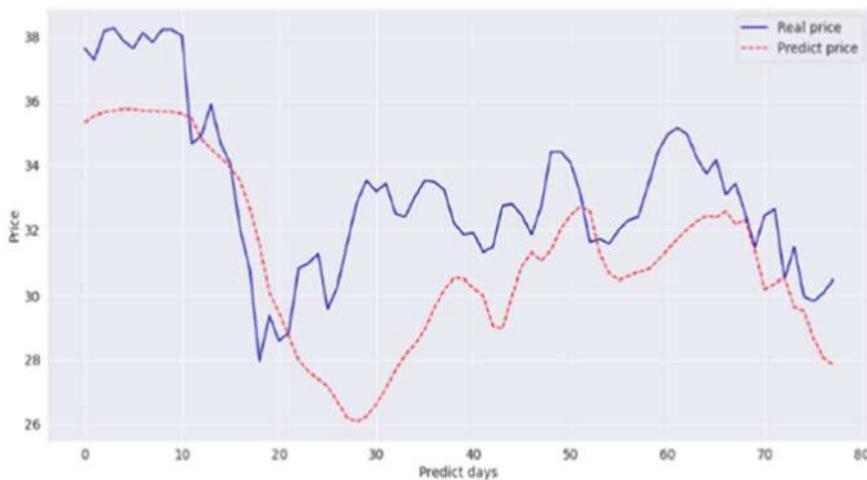
Figure 4 shows the prediction of SPP with respect to the proposed LSTM with AO. Here, the graph is plotted for actual and predicted stock data.

**Table 1** Comparative analysis

Methods	MSE	RMSE	MAE	$R^2$
RNN	0.35	0.98	3.12	0.513
LSTM	0.89	0.891	2.45	0.541
BILSTM	0.03	0.45	1.93	0.621
Proposed	0.01	0.08	1.00	0.783

**Fig. 3** Comparison of various models





**Fig. 4** Prediction of SPP

## 5 Conclusion

This paper introduced a LSTM with AO for SPP and the primary objective of this paper was to identify stock market prices, employing the DJIA dataset. The algorithm AO was employed to dynamically optimize the LSTM model, aiming to achieve optimal results. The proposed method integrates LSTM with AO to enhance stock market price prediction accuracy. By capturing complex temporal patterns through LSTM and efficiently tuning model parameters using AO, this hybrid approach offers a robust solution for forecasting stock prices, aiding investors and analysts in making informed decisions. The MSE and  $R^2$  values achieved were 0.01 and 0.783. Furthermore, the paper suggests that there is room for enhancement in the prediction system in future research. The authors propose exploring adaptive optimizers for improving the model's performance.

## References

1. Kavinnilaa J, Hemalatha E, Jacob MS, Dhanalakshmi R (2021) Stock price prediction based on LSTM deep learning model. In: 2021 international conference on system, computation, automation and networking (ICSCAN). IEEE, pp 1–4
2. Gandhmal DP, Kumar K (2019) Systematic analysis and review of stock market prediction techniques. Comput Sci Rev 34:100190
3. Pang X, Zhou Y, Wang P, Lin W, Chang V (2020) An innovative neural network approach for stock market prediction. J Supercomput 76:2098–2118
4. Mukherjee S, Sadhukhan B, Sarkar N, Roy D, De S (2023) Stock market prediction using deep learning algorithms. CAAI Trans Intell Technol 8(1):82–94

5. Thakkar A, Chaudhari K (2021) Fusion in stock market prediction: a decade survey on the necessity, recent developments, and potential future directions. *Inform Fusion* 65:95–107
6. Ingle V, Deshmukh S (2021) Ensemble deep learning framework for stock market data prediction (EDLF-DP). *Glob Transitions Proc* 2(1):47–66
7. Gürmez B (2023) Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm. *Expert Syst Appl* 227:120346
8. Liu H, Long Z (2020) An improved deep learning model for predicting stock market price time series. *Digital Signal Process* 102:102741
9. Gunduz H (2021) An efficient stock market prediction model using hybrid feature reduction method based on variational autoencoders and recursive feature elimination. *Financ Innov* 7(1):28
10. Rezaei H, Faaljou H, Mansourfar G (2021) Stock price prediction using deep learning and frequency decomposition. *Expert Syst Appl* 169:114332
11. Patil PR, Parasar D, Charhate S (2024) Wrapper-based feature selection and optimization-enabled hybrid deep learning framework for stock market prediction. *Int J Inform Technol Decis Making* 23(01):475–500
12. <https://www.kaggle.com/datasets/mnassrib/dow-jones-industrial-average>

# Studying the Differences in Functioning Among Several Effective Image Steganography Methods



Raghda Salam Al. Mahdawi<sup>ID</sup>, Warqaa Shaher Alazawee<sup>ID</sup>,  
Ali J. Abboud<sup>ID</sup>, and Azmi Shawkat Abdulbaqi<sup>ID</sup>

**Abstract** Nowadays, image steganography is playing a vital role in data security which has increased significantly. By covertly encrypting data bits that have a lower chance of being discovered, image steganography protects data. To date, a number of steganography and cryptographic methods have been created to guarantee the security of data while it is being transmitted over a network. The paper proposes different approaches of Image steganography to secure information in military offices. Also, it is to assess how well various steganography algorithms (LSB, PVD, EMD, PPM, and GLM) work. In this comparison, the benefits and drawbacks of the currently used image steganography methods are highlighted. Metrics are utilized to report the performance of these algorithms in an unbiased manner, including Peak-Signal-to-Noise Ratio (PSNR), Mean Squared Error Measurement (MSE), Structural Similarity Index (SSIM), Noise Visibility Function (NVF), histogram analysis, and standard deviation. In experiments we've done, we discovered that the LSS and PVD algorithms have the best performance out of all the algorithms that were compared, with the EMD and PPM having good performance but less so. In terms of embedding time, PPM and PVD are the most effective algorithms, whereas EMD and GLM are less effective because they take more time. The best and most effective algorithms for recovering data are PVD and PPM, while LSB, EMD, and GLM consume the most resources in terms of recovery time.

**Keywords** Image steganography · Cryptography · Algorithms of steganography

---

R. S. Al. Mahdawi · W. S. Alazawee · A. J. Abboud

Department of Computer Engineering, University of Diyala, Diyala, Baqubah, Iraq  
e-mail: [raghdasalam@uodiyala.edu.iq](mailto:raghdasalam@uodiyala.edu.iq)

W. S. Alazawee

e-mail: [warqaash@uodiyala.edu.iq](mailto:warqaash@uodiyala.edu.iq)

A. J. Abboud

e-mail: [Abboudali.j.abboud@uodiyala.edu.iq](mailto:Abboudali.j.abboud@uodiyala.edu.iq)

A. S. Abdulbaqi (✉)

Renewable Energy Research Center, University of Anbar, Ramadi, Anbar, Iraq  
e-mail: [azmi\\_msc@uoanbar.edu.iq](mailto:azmi_msc@uoanbar.edu.iq)

## 1 Introduction

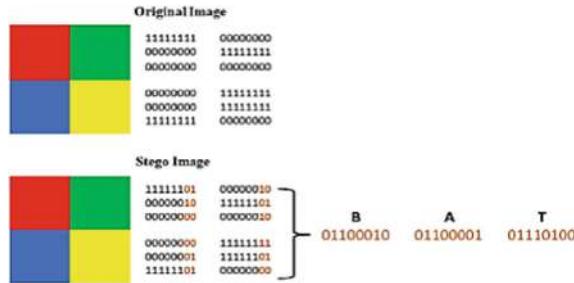
Secret information is scrambled and modified by cryptography so that strangers can't read it [1]. To accomplish cryptography, one can use either standard or Chaos-based methods of encryption. DES, AES, Rivest Shamir Adleman algorithm (RSA), and many others are among the most commonly used encryption standards [2-4]. Data are encrypted before embedding using a private key [2]. Data encryption with SETs is unreliable and insecure because of the limited number of keys [5]. Chaos-based encryption methods have overcome SETs' limitations. Encryption methods based on chaos use initial encryption keys that are susceptible to modifications [3] Therefore, chaos-based encryption methods are under development. Consequently, chaos-based encryption methods provide a more secure cryptographic method for securing data. With encryption, data can be made significantly more secure by changing their shape [4]. Despite its encrypted form, cryptography alone can't resist security threats since attackers are able to manipulate or hack its encrypted form. Although its encrypted form can prevent eavesdroppers from gaining access to the data, it is insufficient when it comes to data security. Researchers have used data hiding extensively to hide from intruders the existence of important information [6].

Steganography and watermarking are the two categories under which data concealment is subdivided [7]. The process of adding a watermark to multimedia files to ensure their legitimacy is known as copyright. The process of adding a watermark to multimedia files to ensure their legitimacy is known as "copyright protection" [8]. To prove ownership, watermarking may be either visible or invisible. Steganography is the practice of obfuscating crucial information in any kind of multimedia for covert communication [9]. Data cannot easily be detected in this procedure since it is invisible. Two types of steganography can be distinguished: Approaches to technical steganography and linguistic steganography [10]. The use of multimedia and technology can further differentiate technical steganography into two categories. The only challenge with steganography technology is maintaining image quality while carrying a significant payload [11].

## 2 Method

### 2.1 Algorithms of Steganography

Information security uses a variety of image steganography algorithms. This chapter briefly describes most of them. Our discussion began with the least significant bit (LSB) and pixel value difference (PVD), which the authors see as the most fundamental. Our next step consists of using additional cutting-edge techniques to find a strategy by increasing embedding efficiency. The following scenario provides an opportunity to discuss different levels of steganography extraction and embedding. The following is a description of comparative steganography algorithms:



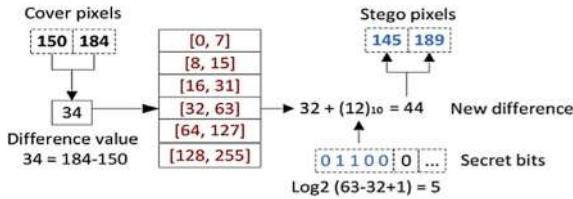
**Fig. 1** Bit plane decomposition of Lena grayscale image

## 2.2 Least Significant Bit (LSB)

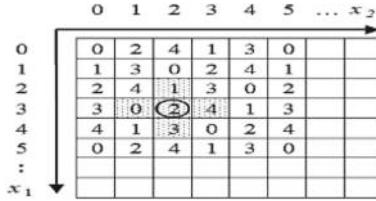
This work's least significant bit (LSB) methodology was spatial domain steganography-in-replacement. Private information is replaced with the least significant bit of a cover image [5]. By removing a specific bit from each pixel that makes up a particular plane of bit, one can represent an 8-bit binary by using each pixel's grayscale value in a 256-bit grayscale cover image, such as the least significant bit of all pixels that make up the least effective bit plane, for example. Gray values contribute differently depending on the bit plane, with the lowest bit plane similar to random noise [12, 13] (Fig. 1).

## 2.3 Pixel Value Differencing (PVD)

Due to the growing and urgent need for data transmission across social networks, information concealment is currently regarded as one of the most essential concerns in human societies [14]. Using cloud services and the Internet, a lot of data is often exchanged through open networks and unsecure channels, exposing private and secret data to harmful situations. As a result, it is essential to make sure that information is transmitted over the Internet in a secure and safe manner [15]. A number of data hiding techniques, including steganography, have been developed to prevent an unauthorized individual from accessing the transmitted information. Steganography is used as the term describing the art of embedding secret messages into some innocuous cover media in such a way that the resultant statistical noise introduced is minimal [16] (Fig. 2).



**Fig. 2** Pixel value differencing steganography



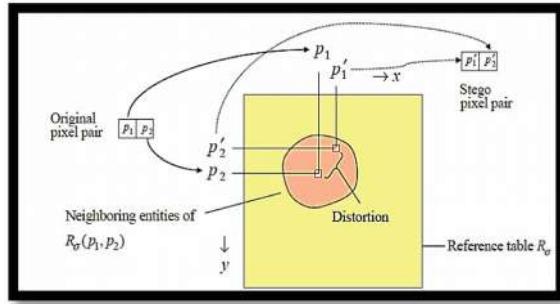
**Fig. 3** Exploiting modification direction steganography

## 2.4 Exploiting Modification Direction (EMD)

In this paper, a new method of information hiding is proposed by using the methodology of EMD and further enhancing it by using the Knight Tour algorithm to embed the information in the image and LZW for encrypting and compressing the secret message [17]. This approach has used LZW for encryption, compression, and bit mapping of the characters of the secret message for its shrinking and protecting [18]. After that, secret numbers have been embedded at places where cover image has been divided into groups using the EMD technique and Knight tour algorithm. There are four pixels per group. The image is further divided into groups of blocks of size (4\*4) for each group.  $n + 1$ -ary theorem [19, 20] (Fig. 3).

## 2.5 Pixel Pair Matching (PPM)

Signal processing has recently experienced a surge of development in the area of steganography. Information is buried in other information to conceal the very idea that communication is even taking place. In this study, it is recommended to hide a variety of data types within images, such as text, images, and music, including Secure Adaptive Pixel Pair Matching (SAPPMM), a method for encrypting both grayscale and color images [21]. Using a secret key supplied by the user, a pseudo-random sequence is generated to encrypt the secret message. A SAPPMM method is used to mask the encrypted message in the cover image. Compared to LSB substitution, Simple Diamond is more efficient [22, 23] (Fig. 4).



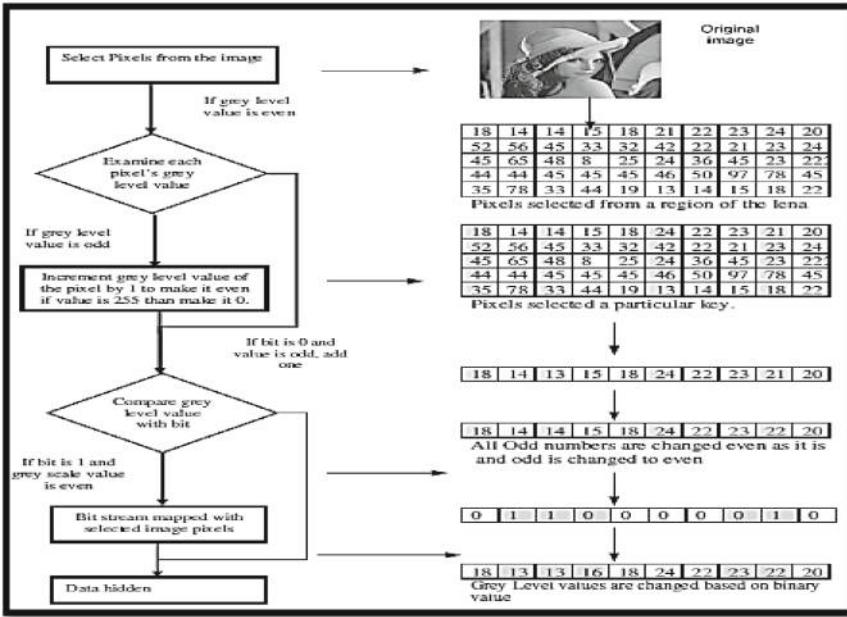
**Fig. 4** PPM-based embedding method is depicted in the following schematic

## 2.6 Gray-Level Modification (GLM)

In the current era of digital technology, steganography is a remarkable instrument. Due to its use, it is quickly acquiring significance. In a short period of time, steganography has garnered a lot of attention [24]. Some have asserted that Al Qaeda used the steganography system to time the World Trade Center attack in their analysis of the events on September 11, 2001 [25]. But nothing was made available for verification after that. The most crucial tool for the safe electronic transmission of sensitive information, document authentication, document tracking, digital elections, and electronic money is just a few of the commercial and scientific uses for hidden writing. In addition to these, the images can be combined with data gathered at a radar station or during medical imaging. A comprehensive system for steganography [26, 27] (Fig. 5).

## 3 Results and Discussion

These studies have shown us that our system is capable of choosing the best algorithm for a particular class of applications. The LSB and EMD are two of the best algorithms for information concealment, whereas the GLM is the approach with the lowest performance, as shown by the PSNR metric values in Fig. 9a. Results from the SSIM measure are shown in Fig. 9b, which demonstrate the superiority of most methods other than the GLM algorithm. The bits per pixel values for each method are then shown in Fig. 9c, which demonstrates the superiority of the PVD algorithm over all other compared algorithms. LSB is the second-best algorithm, followed by PPM, GLM, and EMD. The embedding time, which quantifies the time required to insert data inside a cover image, is another significant metric.



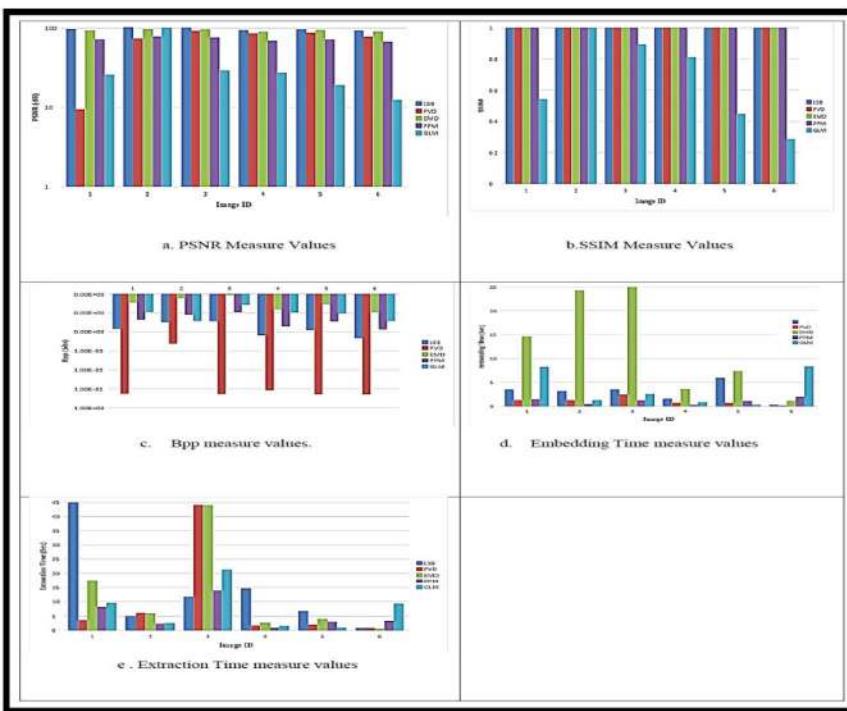
**Fig. 5** Schematic illustration of the gray-level modification (GLM)-based embedding method

Based on Figs. 6, 7, 8, 9, 10, 11, 12, and 13, the figures show examples of applying the compared algorithms (LSB, PVD, EMD, PPM, and GLM) on different images. We conduct experiments by embedding text and images inside cover images and then recover embedded data by extraction algorithms as shown below in these figures. Also, attacks are applied on the cover image to notice whether compared five steganography algorithms can resist these attacks. The results indicate that the compared algorithm is able to resist confronted attacks and get back original embedded text or image inside the cover image with high-quality recovery image (Figs. 14, 15, 16, 17 and 18).

### 3.1 Conclusions

In this project, we have examined the importance of image steganography algorithms. These including LSB, PVD, EMD, PPM, GLM, and MSB algorithms. The best performance algorithm is the PVD and LSB among all compared algorithms. The EMD and PPM have good performance but less than LSS and PVD algorithms. Accurate results were obtained using the Peak-Signal-to-Noise Ratio (PSNR), Mean

Squared Error Measurement (MSE), Structural Similarity Index (SSIM), Noise Visibility Function (NVF), histogram analysis, and standard deviation analysis. PPM and PVD are the most efficient algorithms in terms of embedding time, while EMD and GLM are the less efficient algorithms because they consume more time. PVD and PPM are the best and efficient algorithms in recovering data, while the LSB, EMD, and GLM are the highest algorithms in terms of consuming recovery time resource. Our work will expand to use other steganography algorithms for other types of applications in future. These challenges associated with this science are the large volume of these image data and their special properties. Also, we can use machine learning algorithms to improve the performance of these algorithms.



**Fig. 6** Values of applying various (LSB, PVD, PPM, EMD, and GLM) on different images

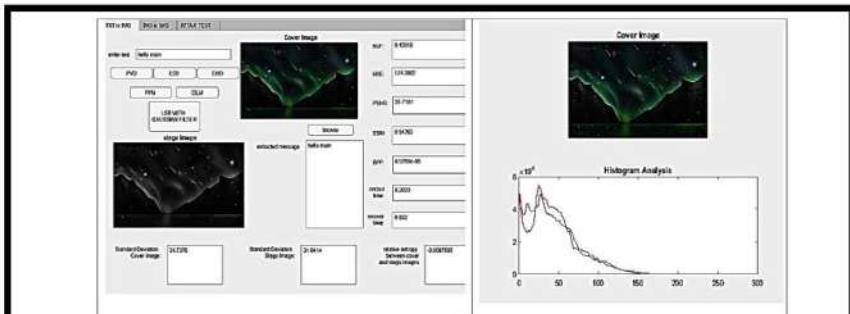


Figure 6(a,b). example one

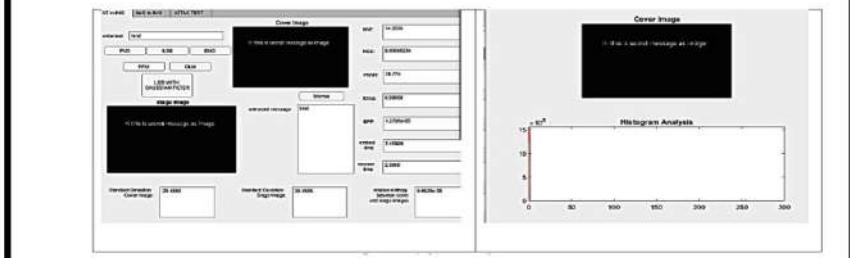
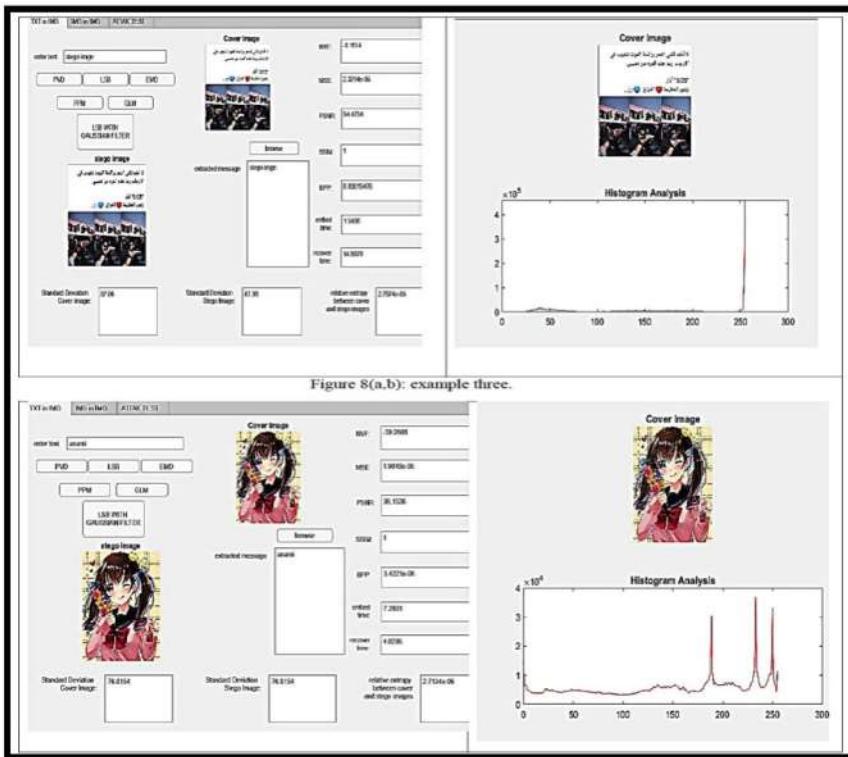
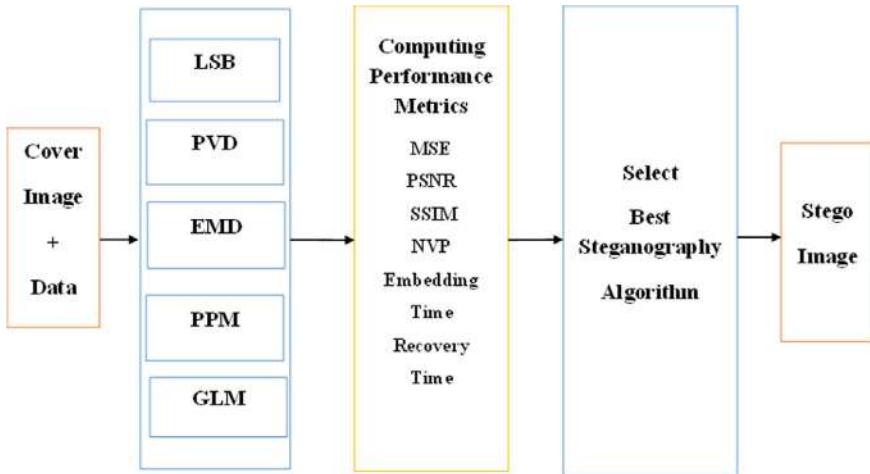


figure 7 (a,b) example two

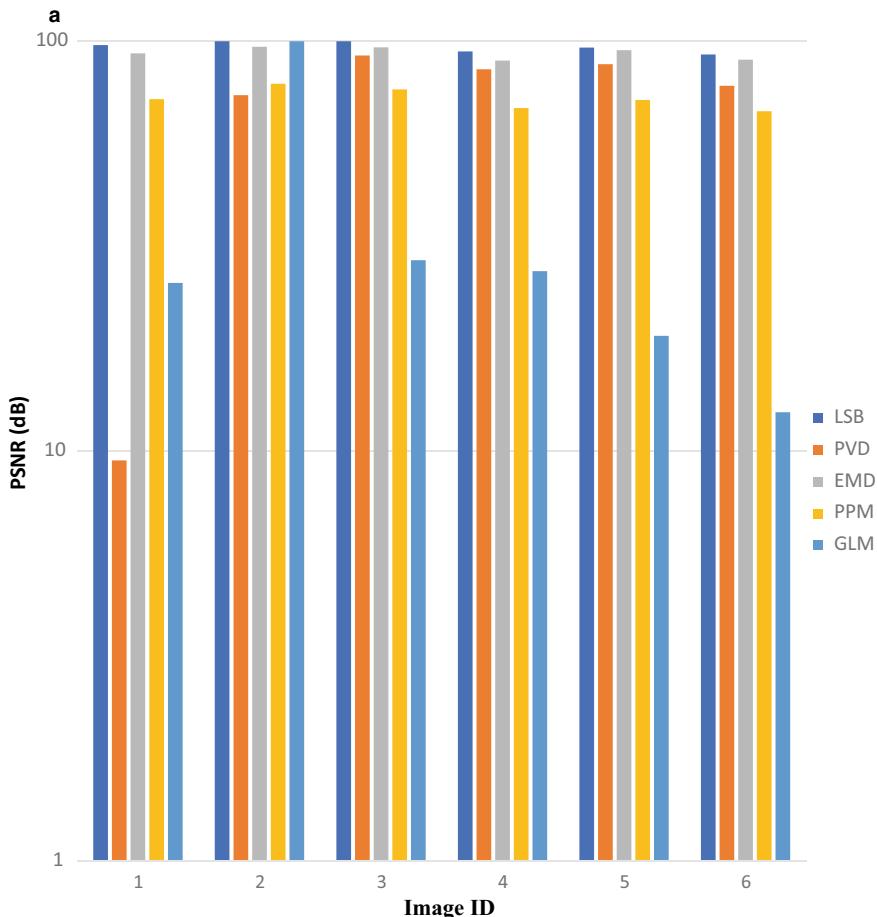
**Fig. 7** Both examples 1 and 2 applied on the proposed methods



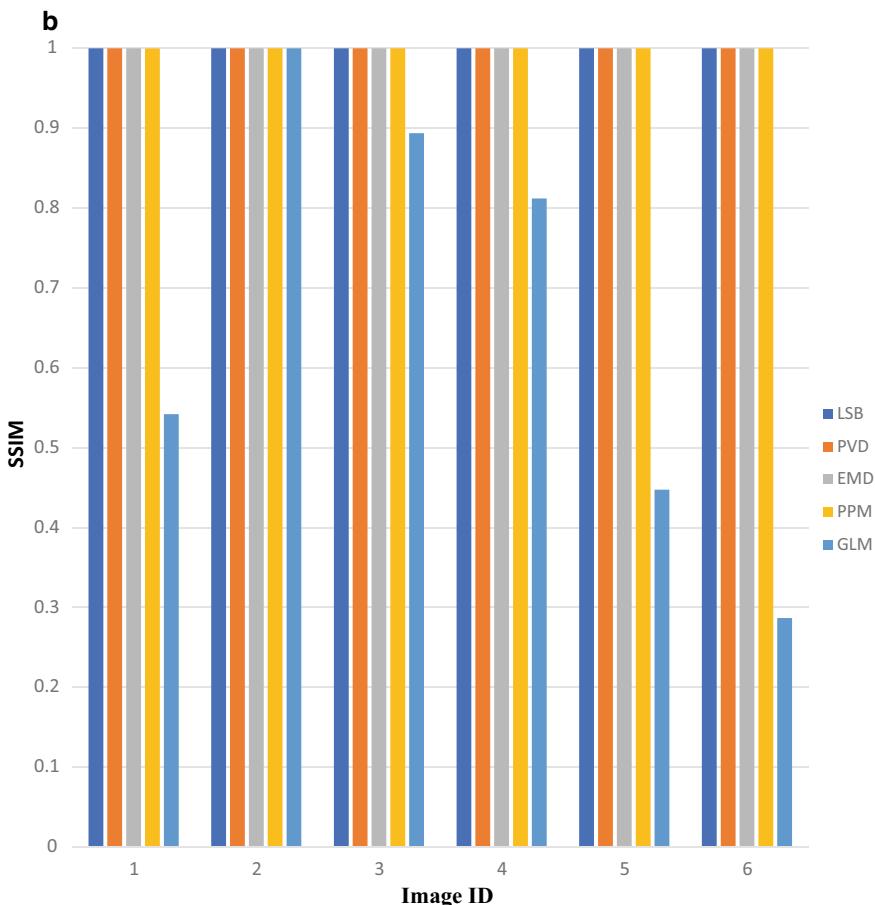
**Fig. 8** Both example 3 and 4 applied on the proposed methos



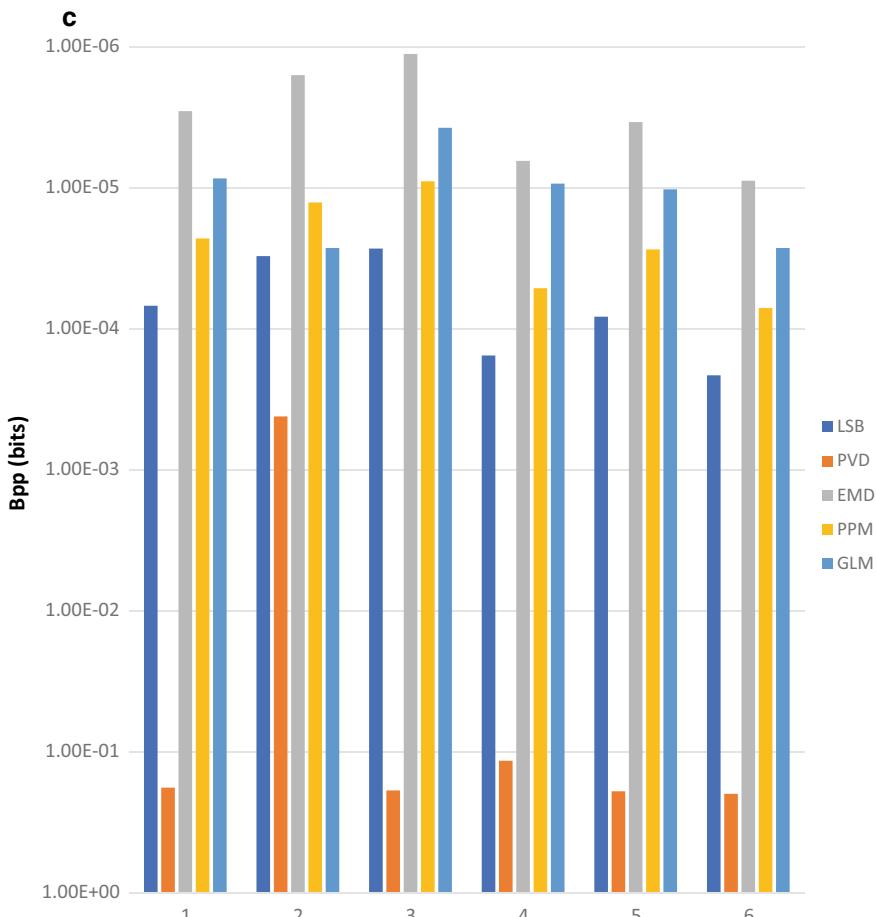
**Fig. 9** Block diagram of measurement tool to select best steganography algorithm



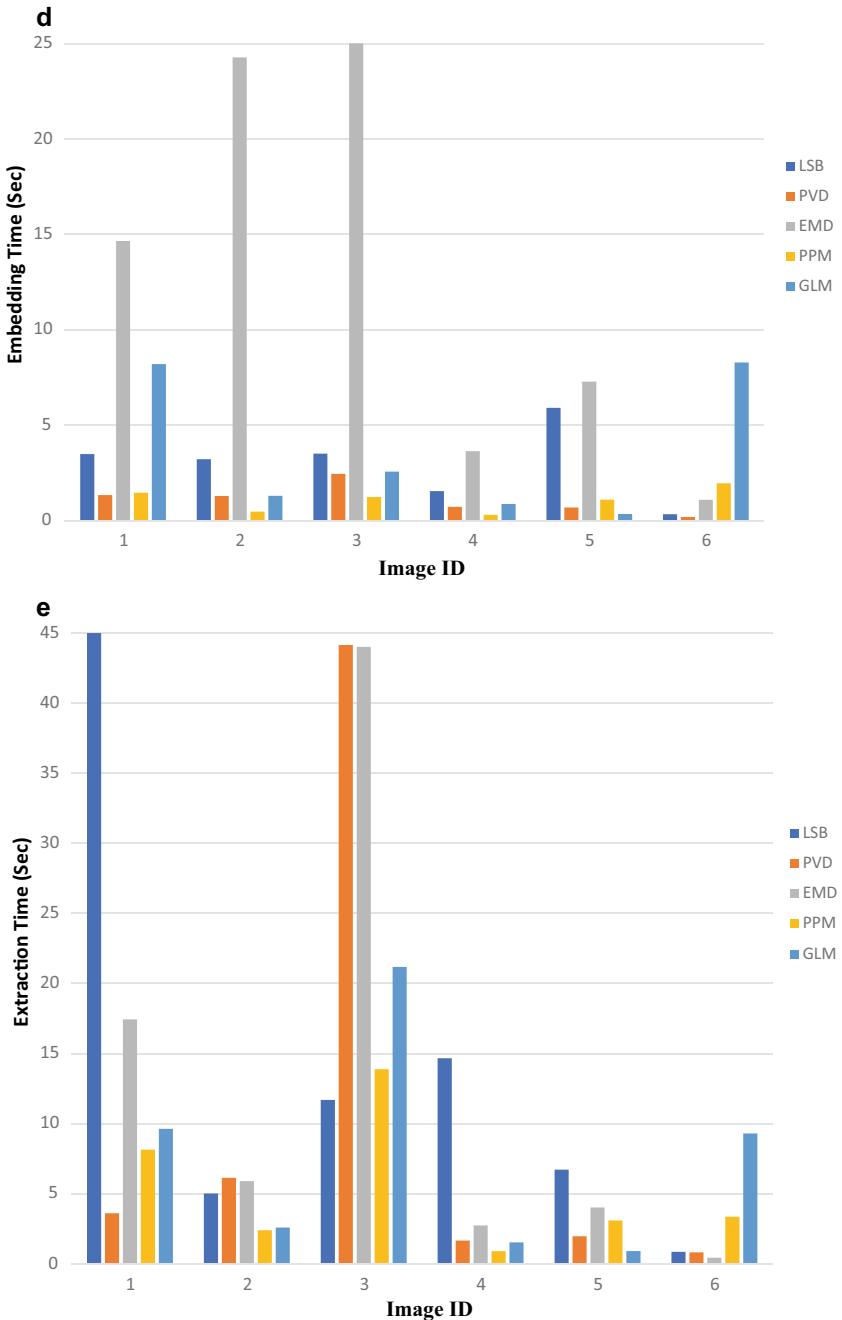
**Fig. 10** Results of performance evaluation of different steganography algorithms. **a** PSNR measure values **b** SSIM measure values **c** Bpp measure values **d** Embedding time measure values **e** Extracition time measure values



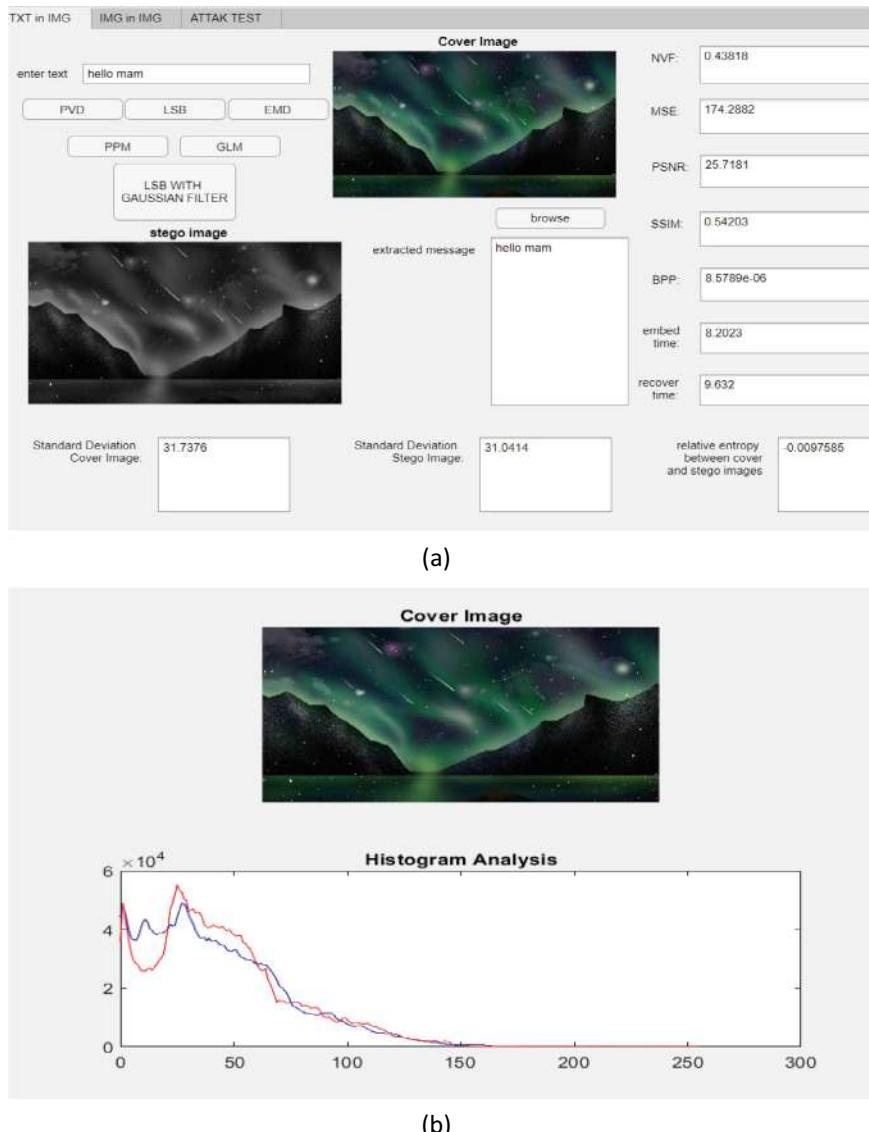
**Fig. 10** (continued)

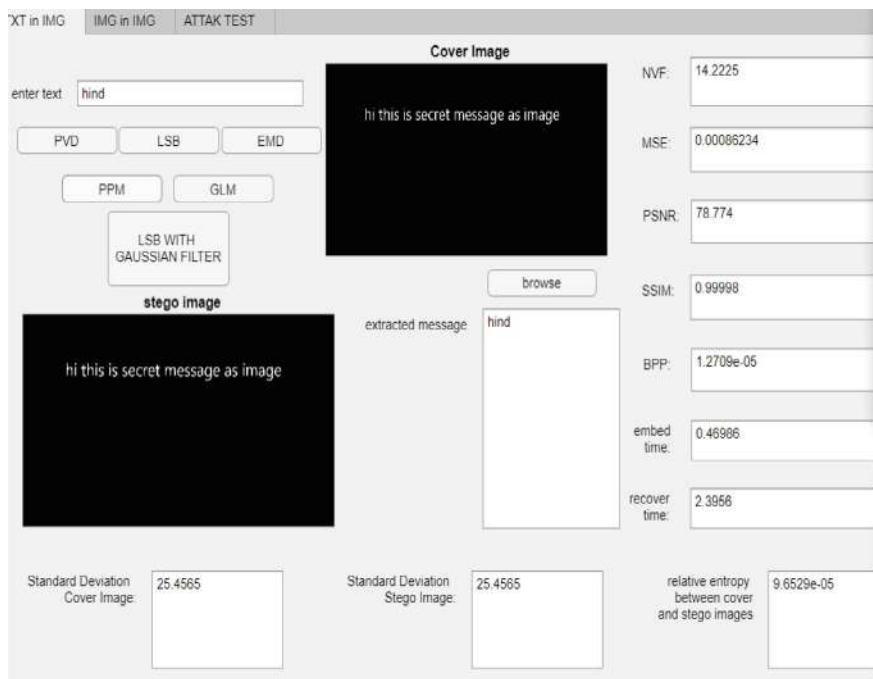


**Fig. 10** (continued)

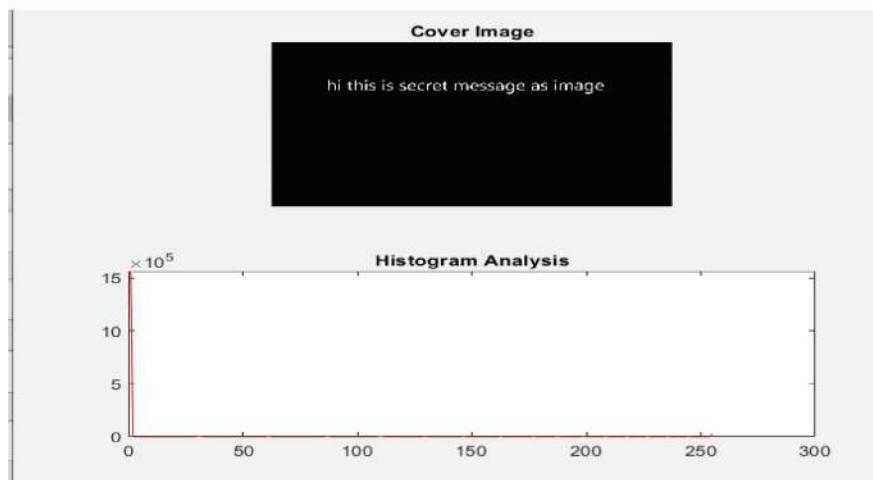


**Fig. 10** (continued)

**Fig. 11** Result of the GLM algorithm

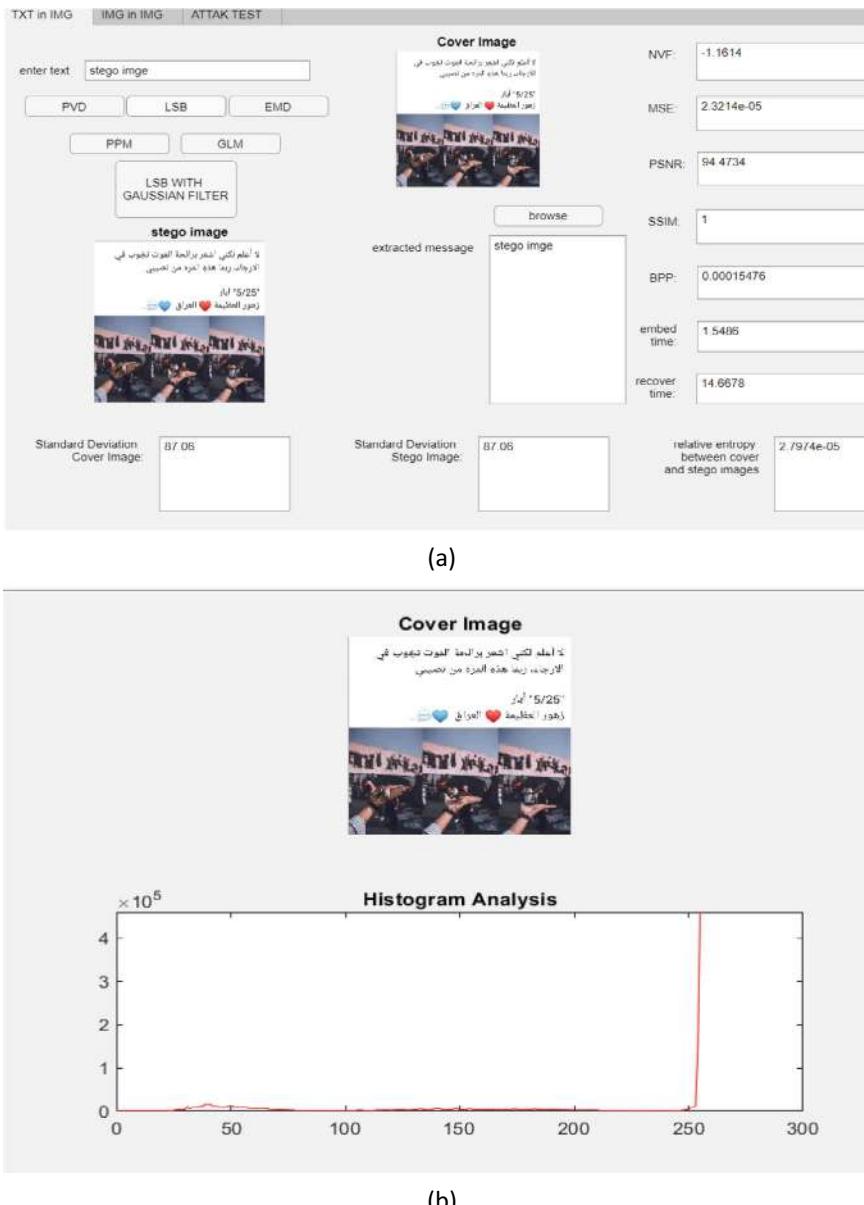


(a)



(b)

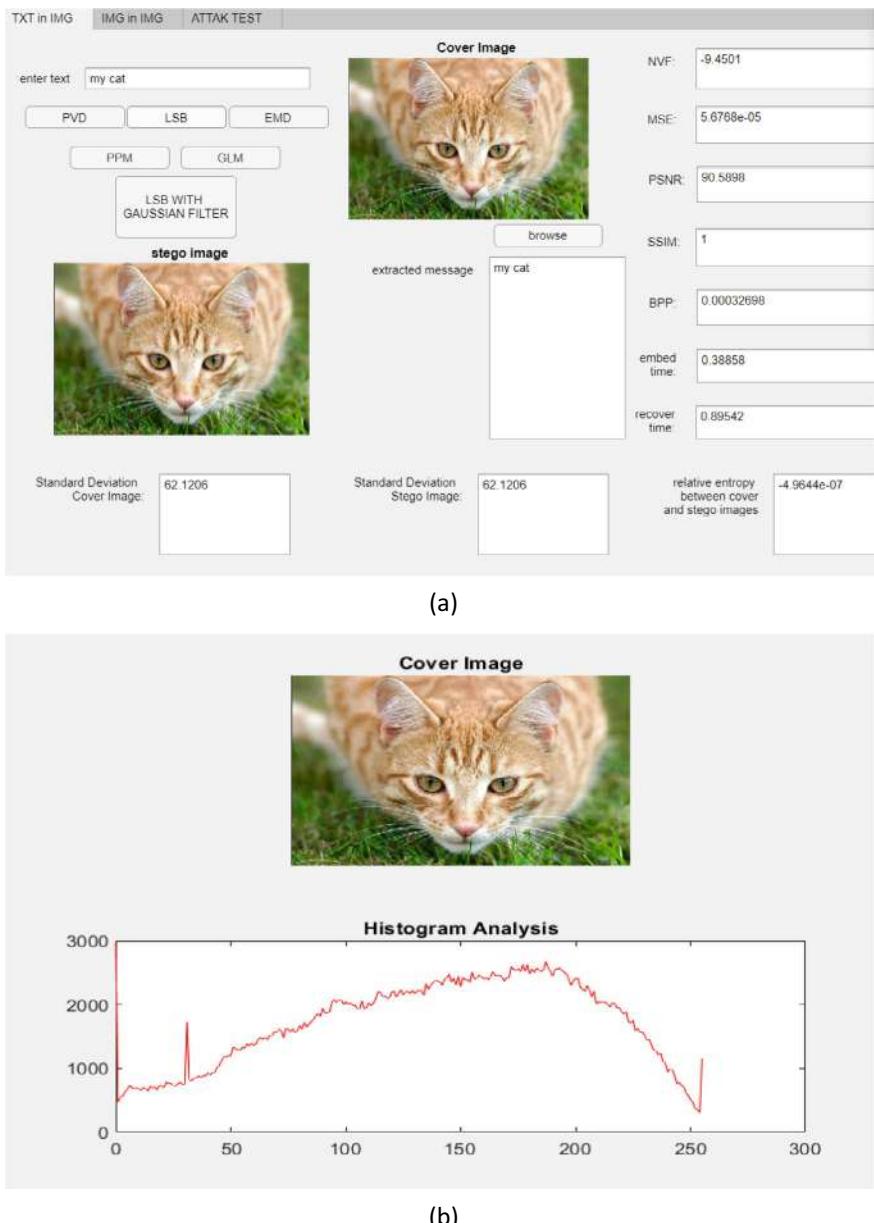
**Fig. 12** Result of the LBS algorithm

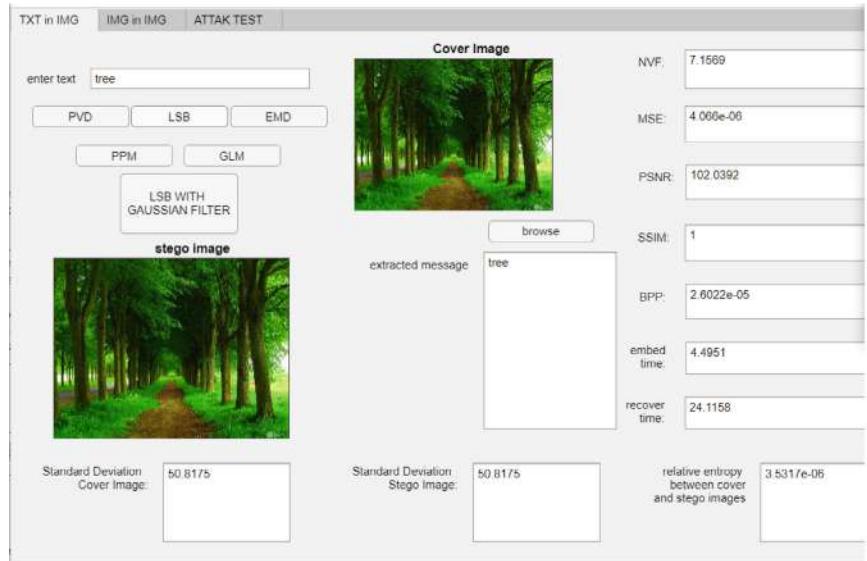


**Fig. 13** EMD algorithm result

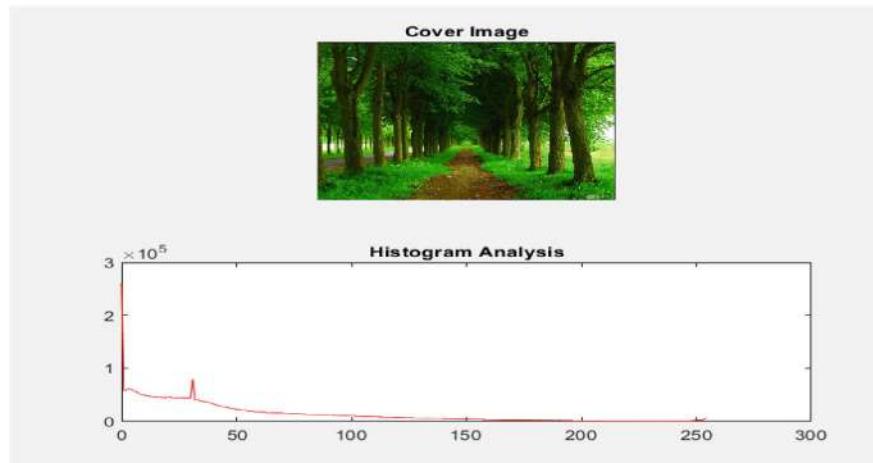


**Fig. 14** PPM algorithm result

**Fig. 15** LBS algorithm result

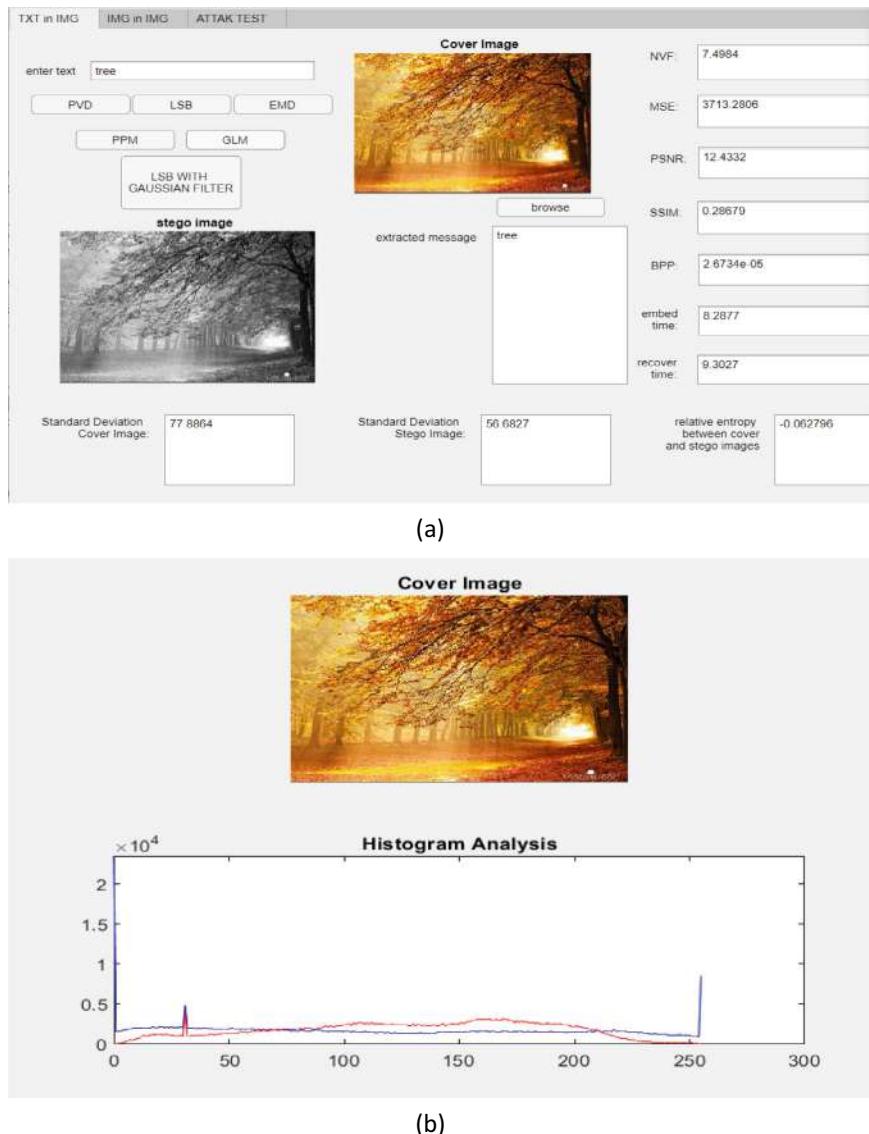


(a)



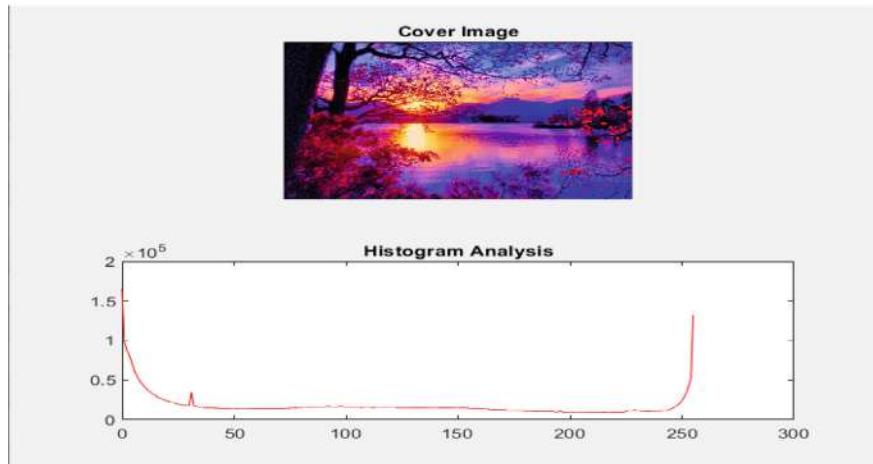
(b)

**Fig. 16** PVD algorithm result

**Fig. 17** Result of the GLM algorithm



(a)



(b)

**Fig. 18** PVD algorithm result

## References

1. Khan M, Sajjad M, Mehmood I, Rho S, Baik SW (2016) Image steganography using uncorrelated color space and its application for security of visual contents in online social networks. *Futur Gener Comput Syst*
2. Nguyen G et al (2019) Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey. *Artif Intell Rev* 52(1):77–124. <https://doi.org/10.1007/s10462-018-09679-z>
3. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. *J Big Data* 6(1). <https://doi.org/10.1186/s40537-019-0197-0>
4. Vinayakumar R, Alazab M, Soman KP, Poornachandran P, Al-Nemrat A, Venkatraman S (2019) Deep learning approach for intelligent intrusion detection system. *IEEE Access* 7:41525–41550. <https://doi.org/10.1109/ACCESS.2019.2895334>
5. Lau BPL et al (2019) A survey of data fusion in smart city applications. *Inf Fusion* 52(January):357–374. <https://doi.org/10.1016/j.inffus.2019.05.004>
6. Al Mahdawi RS, Salih HM (2021) Optimization of open flow controller placement in software defined networks. *Int J Electr Comput Eng* 11(4):3145–3153
7. Jayalakshmi G, Khalaf HA, Farhadi A, Mahdawi RSA, Abdulbaqi AS (2022) Detection of COVID-19 from radiology modalities and identification of prognosis patterns. *Int J Nonlinear Anal Appl* 13(1):1351–1365
8. Singh AK, Kumar B, Singh SK, Ghrera SP, Mohan A (2016) Multiple watermarking technique for securing online social network contents using back propagation neural network. *Futur Gener Comput Syst*
9. Kumar S, Singh M (2019) Big data analytics for healthcare industry: impact, applications, and tools. *Big Data Min Anal* 2(1):48–57. <https://doi.org/10.26599/BDMA.2018.9020031>
10. Rahman NSA, Rahim NA (2023) Sustainable framework for a geostationary satellite control earth station system using parallel configuration. *Indonesian J Electr Eng Comput Sci* 30(3)
11. Abdulbaqi AS, Obaid AJ, Mohammed A (2021) ECG signals recruitment to implement a new technique for medical image encryption. *J Discrete Math Sci Crypt* 24(6):1663–1673. <https://doi.org/10.1080/09720529.2021.1884378>
12. Marakumbi P, Bhairannawar S (2023) Efficient reconfigurable architecture to enhance medical image security. *Indonesian J Electr Eng Comput Sci* 30(3)
13. Mosavi A, Shamshirband S, Salwana E, Chau KW, Tah JHM (2019) Prediction of multi-inputs bubble column reactor using a novel hybrid model of computational fluid dynamics and machine learning. *Eng Appl Comput Fluid Mech* 13(1):482–492. <https://doi.org/10.1080/19942060.2019.1613448>
14. Palanisamy V, Thirunavukarasu R (2019) Implications of big data analytics in developing healthcare frameworks—a review. *J King Saud Univ—Comput Inf Sci* 31(4):415–425. <https://doi.org/10.1016/j.jksuci.2017.12.007>
15. Sadowski J (2019) When data is capital: datafication, accumulation, and extraction. *Big Data Soc* 6(1):1–12. <https://doi.org/10.1177/2053951718820549>
16. Saura JR, Herraez BR, Reyes-Menendez A (2019) Comparing a traditional approach for financial brand communication analysis with a big data analytics technique. *IEEE Access* 7:37100–37108. <https://doi.org/10.1109/ACCESS.2019.2905301>
17. Nallaperuma D et al (2019) Online incremental machine learning platform for big data-driven smart traffic management. *IEEE Trans Intell Transp Syst* 20(12):4679–4690. <https://doi.org/10.1109/TITS.2019.2924883>
18. Schulz S, Becker M, Groseclose MR, Schadt S, Hopf C (2019) Advanced MALDI mass spectrometry imaging in pharmaceutical research and drug development. *Curr Opin Biotechnol* 55:51–59. <https://doi.org/10.1016/j.copbio.2018.08.003>
19. Shang C, You F (2019) Data analytics and machine learning for smart process manufacturing: recent advances and perspectives in the big data era. *Engineering* 5(6):1010–1016. <https://doi.org/10.1016/j.eng.2019.01.019>

20. Song Q, Ge H, Caverlee J, Hu X (2017) Tensor completion algorithms in big data analytics. 13(1) arXiv
21. Yu Y, Li M, Liu L, Li Y, Wang J (2019) Clinical big data and deep learning: applications, challenges, and future outlooks. *Big Data Min Anal* 2(4):288–305. <https://doi.org/10.26599/BDMA.2019.9020007>
22. Huang M, Liu W, Wang T, Song H, Li X, Liu A (2019) A queuing delay utilization scheme for on-path service aggregation in services-oriented computing networks. *IEEE Access* 7:23816–23833. <https://doi.org/10.1109/ACCESS.2019.2899402>
23. Xu G, Shi Y, Sun X, Shen W (2019) Internet of things in marine environment monitoring: a review. *Sens (Switz)* 19(7):1–21. <https://doi.org/10.3390/s19071711>
24. Aqib M, Mehmood R, Alzahrani A, Katib I, Albeshri A, Altowaijri SM (2019) Smarter traffic prediction using big data, in-memory computing, deep learning and GPUs 19(9)
25. Stylos N, Zwiegelaar J (2019) Big data as a game changer: how does it shape business intelligence within a tourism and hospitality industry context?
26. Penubadi HR, Shah P, Sekhar R, Alrasheedy MN, Niu Y, Radhi AD, Tharwat M, Tawfeq JF, Gheni HM, Abdulbaqi AS (2023) Sustainable electronic document security: a comprehensive framework integrating encryption, digital signature and watermarking algorithms. *Heritage Sustain Dev* 5(2):391–404. <https://doi.org/10.37868/hsd.v5i2.359>
27. Leonelli S, Tempini N (2020) Data journeys in the sciences

# Monitoring and Optimization of Machine Learning Workloads Using Kubernetes



Ananth Mahesh Kashyap, V. Dinesh Reddy, and Marco Aiello

**Abstract** The demand for energy in cloud-native applications has increased considerably in recent years. With the rise of container-based deployments for delivering applications, understanding their power usage patterns is critical to lowering them. Unfortunately, cloud vendors do not provide their clients with power consumption details for individual workloads owing to virtualization-related limits inside the cloud infrastructure. This research paper compares the software and hardware-based tools available in the market to measure power consumption and discusses in detail about Kubernetes Efficient Power Level Exporter (Kepler), which addresses the above issue by estimating power metrics at the container level by using extended Berkeley Packet Filter (eBPF) and machine learning (ML) models. Since data-intensive workloads are power-hungry, we run the ML models on a simulated Graphical Processing Unit (GPU) accelerated Kubernetes (K8s) cluster. The metrics extracted by Kepler are carefully analyzed, and the ML workloads are tuned and optimized to use less energy.

**Keywords** Kubernetes · Kubernetes Efficient Power Level Exporter · Extended Berkeley packet filter · Machine learning · Cloud · Virtualization

## 1 Introduction

The rise in popularity of containers has brought about a major shift in application development and deployment within the contemporary cloud computing landscape. Containers are becoming more and more popular because they are able to solve the

---

A. M. Kashyap  
University of Stuttgart, Stuttgart, Germany

V. Dinesh Reddy · M. Aiello  
Service Computing, IAAS, University of Stuttgart, Stuttgart, Germany  
e-mail: [marco.aiello@iaas.uni-stuttgart.de](mailto:marco.aiello@iaas.uni-stuttgart.de)

V. Dinesh Reddy (✉)  
CSE, SRM University, Amaravati, AP, Germany  
e-mail: [dinesh.vemula@iaas.uni-stuttgart.de](mailto:dinesh.vemula@iaas.uni-stuttgart.de)

“it works on my machine” issue by offering a consistent environment for development, testing, and deployment. However, this development resulted in a notable rise in data center power usage, which now accounts for 2% of global energy consumption. According to recent projections, this percentage is expected to rise even higher, reaching at least 8% by 2030 [8]. Due to several variables, including the acceleration of data collecting, computationally demanding workloads involving Artificial Intelligence (AI), and the slowing down of Moore’s law, the amount of power used in cloud computing is rapidly increasing. The increasing energy demand in cloud infrastructures has raised worries about its long-term sustainability and ecological effects.

Cloud providers typically monitor the aggregate power consumption of the infrastructure at the data center level. However, they lack detailed power consumption metrics of individual workloads or containers running these workloads. This lack of granularity does not allow the customers to optimize their applications to be more efficient and consume less energy. Furthermore, it is difficult to detect and reduce operations that consume a lot of power without exact data, which results in inefficiencies and increased operational expenses. Implementing targeted energy-saving methods is made more difficult by the lack of workload-specific power metrics, which is problematic given the increasing environmental concerns and rising energy costs.

This paper recognizes the gap in power consumption extraction and utilizes the Kubernetes-based Kepler framework, which collects the data at the container and pod levels and uses power models to estimate the real-time power usage of operations. Power consumption may be extracted down to the container level, which allows for power optimization of both the compute nodes and the applications. This detailed method promotes dynamic adjustments based on real-time data, which improves overall energy efficiency. The most popular container orchestration technology, Kubernetes, is being employed as the container orchestration framework since container-based workloads are of interest in this study.

This paper is thoughtfully structured across several sections where Sect. 2 presents a comprehensive comparison of the tools available in the market that use both hardware and software to measure power consumption. Section 3 defines the architectural configuration and the hardware and software requirements to deploy Kepler on a Minikube cluster and power measurement during the training of ML workloads. Section 4 describes the experiments conducted with two different neural networks. This section also assesses and seeks to uncover patterns in the power consumption metrics for the ML models, which might aid in optimizing the workload to consume less power. Finally, Sect. 5 concludes the paper by summarizing the findings, discussing their implications, and providing recommendations for future work.

## 2 Related Works

This section dives into some of the research articles that provide a full assessment of the hardware and software-based techniques used to measure workload power consumption. A thorough comparison of the tools is conducted, which assisted in selecting Kepler as the tool for measuring power at the container level.

The study by Jay et al. in [2] compares many software-based power meters designed for CPU or GPU-based infrastructures both experimentally and qualitatively. The most established and least invasive method of estimating a computing node's power consumption is to use physical power meters. However, they necessitate the deployment of additional measuring infrastructure. A physical power meter merely tracks the total usage of the computing node. This does not include the consumption of any components or services launched on the compute node. This research classifies software-based power meters into three categories:

- **Energy Calculators** are web-based software power meters that employ Thermal Design Power (TDP) utilization modeling to calculate device energy consumption [2].
- **Energy Measurement Software** packages calculate the total energy expended by the compute node during program execution, and Python software packages were chosen for their compatibility with Intel RAPL and Nvidia NVML [2].
- **Power Profiling Software** tools are studied to provide insights into program power consumption trends, using the Intel RAPL and/or Nvidia NVML interfaces [2].

The software-based tools mentioned above lack in extracting power consumption at the granular or the container level which motivated us to use the Kepler framework for energy estimation.

Centofanti et al. [8] discuss the critical requirement to monitor energy consumption in container-based solutions, which have become crucial to modern cloud-native applications. The study presents a complete overview of methods for measuring power consumption across common cloud platforms. Stress terminal UI (s-tui) is an open-source software program that lets you stress test and monitor system performance via a terminal or command-line interface (CLI) [12]. The Kubernetes Efficient Power Level Exporter (Kepler) aims to collect power consumption measurements across multiple platforms to better understand power consumption in container cloud architectures [1]. Scaphandre is a power monitoring program that uses system-level data and hardware-based power sensors to assess power usage on computing systems. It analyzes resource utilization measurements, such as CPU and memory usage, to link resource activity with power consumption [10]. The Meross MSS310 smart plugs are internet-connected gadgets that give consumers with insights on their current power consumption [11]. These devices are straightforward, consisting of a male-to-female power connector, a button, and a microcontroller equipped with a wireless antenna capable of connecting to a basic Wireless Local Area Network (WLAN) with internet access.

The authors of the paper “Monitoring the Energy Consumption of Docker Containers” discuss about extracting energy consumption of docker containers [3]. In order to highlight the necessity for energy-efficient container development, this article investigates the energy consumption of Docker containers under various workload scenarios. Proposing a way to quantify the energy consumption of Docker containers, examining the variables that influence their energy use, and encouraging advancements in energy efficiency are some of the major contributions. The study provides an in-depth examination of contributing factors, shows results from energy consumption studies, and describes the experimental setting. The intention is to support energy-efficient deployments and give deployment managers practical information. The experiments conducted in this paper are on a single host and cannot be replicated to other platforms which makes it less compatible on cloud-based data centers.

Amaral et al. [1] present the Kepler framework, which aims to assess power usage at many levels, including Kubernetes pod, process, and container. The work suggests a generic power model that uses running average power limit (RAPL) and hardware counters (HC) as independent variables in a regression model that measures system power in real time. Kepler collects the data at the container and pod levels and uses power models to estimate the real-time power usage of operations. The modeling approach builds upon commonly available system power interfaces, such as the built-in Berkeley Packet Filter-based application for capturing in-kernel hardware performance counter events at the context switch level, advanced configuration, and power management interface (ACPI) sensors for measuring system power consumption, and RAPL for gathering CPU and DRAM power consumption.

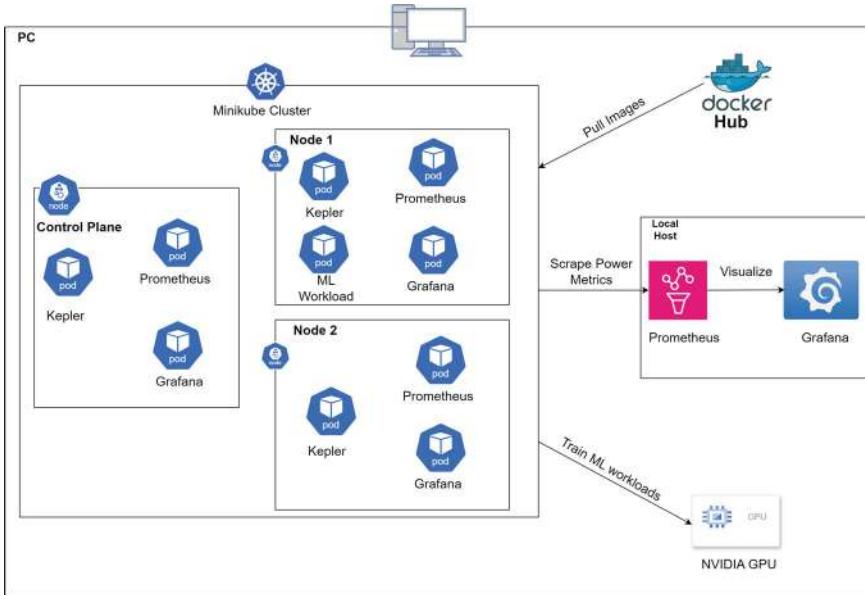
### 3 Proposed Method

This section focuses on the practical implementation, emphasizing collecting power consumption metrics using Kepler that uses eBPF programs and power models to accurately estimate power consumption at the node and pod level of a Kubernetes cluster. The details of the architectural setup, the hardware and software requirements, and the setup of the K8s cluster to exploit the GPU resources available on the system are explained below.

#### 3.1 Architecture

Figure 1 gives a brief overview of the architectural design employed as a part of the implementation.

The implemented architecture, shown in Fig. 1, runs in an Ubuntu environment. Initially, a Minikube K8s Cluster is created, which includes a master node and two slave nodes. Prometheus and Grafana are then installed as Kubernetes pods to help gather and visualize the power measurements retrieved by Kepler. Kepler, which is a



**Fig. 1** Architecture of the proposed system for measuring pod level energy consumption

**Table 1** Hardware and software requirements

Hardware	Software
CPU: 2 cores or more	Docker
GPU: any GPU CUDA cores	Kubectl
Disk space: 50 GB or more	Open lens IDE
Operating system: Ubuntu (v20 or later)	Minikube

power level exporter, is then deployed as a pod on each node in the cluster. Following this configuration, machine learning models are dockerized and deployed to the cluster via kubectl manifest. The Kubernetes cluster trains the model using the available GPU resources, with Prometheus providing continual monitoring of workload power consumption. Through the subsequent examination of these parameters, the workload and the cluster can be optimized, leading to increased resource usage efficiency and decreased energy consumption. The hardware and software requirements for these experiments are listed in Table 1.

The Minikube Cluster by default does not recognize the GPU resource available on the host to deploy the ML workloads. This section outlines the steps taken to make the K8s Cluster aware of the GPU node.

- 1. NVIDIA CUDA Toolkit:** CUDA is a programming methodology and platform for parallel computing. By utilizing the graphics processing unit's (GPU) power,

it allows for a significant gain in computing performance. CUDA-capable GPUs can support thousands of computing threads due to their hundreds of cores [6].

2. **NVIDIA Device Plugin for K8s:** The NVIDIA device plugin is a Daemonset which allows you to:

- Display the quantity of GPUs installed on each cluster node [7].
- Monitor your GPUs' health [7].
- Use containers with GPU support in your Kubernetes cluster [7].

3. **NVIDIA Container Toolkit:** GPU-accelerated container development and execution are made possible by the NVIDIA Container Toolkit. A container runtime library and tools for automatically configuring containers to take advantage of NVIDIA GPUs are included in the toolkit [5].

## 4 Experimental Analysis

In this section, we look at the metrics extracted by Kepler during the training of two ML workloads. The power consumption metrics during the training of the models are analyzed for multiple combinations from a set of hyperparameters. The goal is to find patterns in the power metrics which can lead to less power consumption of the model. The experiments consider two hyperparameters: batch size and learning rate. The training is iterated across nine distinct combinations of batch size and learning rate, with an epoch of 100 serving as the benchmark for training the models [9].

### 4.1 Monitoring ML Workloads Using Kepler

We analyze power consumption metrics captured by Kepler during the training of two machine learning models. By varying batch size and learning rate across different combinations, we aim to identify patterns that could reduce power usage.

**Usecase 1:** The convolutional neural network (CNN) model is utilized as a part of the first experiment because the dataset used in this model is the MNIST dataset [13]. The CNN model is an excellent fit since it deals with data with a grid-like topology, such as photographs. They are ideal for image recognition and categorization. The model consists of two convolutional layers: The first layer has one input and thirty-two output channels, and the second layer has thirty-two input and one output channel. The dataset includes 60,000 training and 10,000 test samples.

**Usecase 2:** The second experiment employs the Siamese Network Model, which utilizes a modified version of the MNIST dataset known as APP MATCHER. The architecture of the siamese neural network is intended to detect the similarity between two inputs [14]. It consists of two identical sub-networks which are CNN networks,

each processing one of the inputs. The model encapsulates a bunch of hidden layers which are: feature extraction layer, concatenation layer, comparison layers, and the output layer.

## 4.2 Results

The power consumption metrics of the GPU for the CNN model and Siamese Network model during training with various combinations of Batch Size and Learning Rate from the above experiments are summarized in Tables 2 and 3. Upon close inspection of both tables, we can infer that the batch size has a higher impact on the model's power consumption, as both CNN and Siamese models consumed less power for larger batch sizes. It can also be observed that the learning rate does not have a major impact on the energy consumption per epoch, the reason being is that the learning rate is independent of the training duration.

The research conducted by Geißler et al. [4] showcases similar results where the energy consumption of the ML workloads decreases with increase in batch size and remains constant across learning rates.

**Table 2** Average power consumption of CNN across all combinations of learning rate and batch size

Batch size	Learning rate		
	0.1	0.01	0.001
64	39.7 (W)	40.1 (W)	40.6 (W)
256	39 (W)	38.1 (W)	38.5 (W)
1024	34.8 (W)	32.2 (W)	30.6 (W)

**Table 3** Average power consumption of siamese model across all combinations of learning rate and batch size

Batch size	Learning rate		
	0.1	0.01	0.001
64	52.4 (W)	52.9 (W)	52.6 (W)
256	52.1 (W)	51.2 (W)	52.3 (W)
1024	51.3 (W)	50.9 (W)	50.6 (W)

## 5 Conclusion

In summary, the application of Kepler in this study advances the fields of data center power monitoring and container-based workload optimization. Through a thorough evaluation of current hardware and software technologies Sect. 2 , we were able to pick Kepler and create a strong framework for tracking Kubernetes workload power consumption. The design that is suggested in Sect. 3 successfully incorporates Kepler into a GPU-supported virtualized Kubernetes cluster, exporting crucial metrics to Grafana for examination. Based on practical testing, we found that while different learning rates had little effect on power usage, larger batch sizes during CNN and Siamese network model training did (Tables 2 and 3). These results demonstrate how well Kepler works to optimize ML model training for energy efficiency.

We plan to extend the proposed architecture to the cloud for individual workload monitoring in our future work. Firstly, the Kubernetes cluster can be deployed utilizing the cloud vendors managed Kubernetes service. The deployment of Kepler is comparable to our recommended solution. It is deployed on every node from which we obtain power metrics. In a cloud environment, the solution can be extended to optimize the cluster by using energy-based scheduling, in which K8s pods are placed on energy-efficient nodes which ensures proper resource utilization, ensuring optimization from the cluster and the workload.

## References

1. Amaral M, Chen H, Chiba T, Nakazawa R, Choochotkaew S, Lee E, Eilam T (2023) Kepler: a framework to calculate the energy consumption of containerized applications. In: 2023 IEEE 16th international conference on cloud computing (CLOUD)
2. Jay M, Ostapenco V, Lefévre L, Trystram D, Orgerie A-C et al (2023) An experimental comparison of software-based power meters: focus on CPU and GPU. In: 23rd IEEE/ACM international symposium on cluster, cloud and internet computing (CCGrid), Bangalore, India, May 2023, pp 1–13. <https://doi.org/10.1109/CCGrid57682.2023.00020>, <https://hal.archives-ouvertes.fr/hal-04030223v2>
3. Warade M, Lee K, Ranaweera C, Schneider J (2023) Monitoring the energy consumption of docker containers. In: 2023 IEEE 47th annual computers, software, and applications conference (COMPSAC), pp 1703–1710. <https://doi.org/10.1109/COMPSAC57700.2023.00263>
4. Geißler D, Zhou B, Liu M, Suh S, Lukowicz P (2024) The power of training: how different neural network setups influence the energy demand. arXiv, 2401.01851. <https://arxiv.org/abs/2401.01851v1>
5. NVIDIA Container Toolkit Documentation. <https://docs.nvidia.com/datacenter/cloud-native/container-toolkit/latest/index.html> Accessed: 3 May 2024
6. NVIDIA Corporation (2022) NVIDIA CUDA installation guide for linux. <https://docs.nvidia.com/cuda/cuda-installation-guide-linux/contents.html>. Accessed: 2 May 2024
7. NVIDIA/k8s-device-plugin: NVIDIA Kubernetes Device Plugin. <https://github.com/NVIDIA/k8s-device-plugin>. Accessed: 3 May 2024
8. Centofanti C, Santos J, Gudepu V, Kondepudi K (2024) Impact of power consumption in containerized clouds: a comprehensive analysis of open-source power measurement tools
9. Understanding the impact of learning rate and batch size in machine learning. <https://www.baeldung.com/cs/learning-rate-batch-size>. Accessed: 7 May 2024

10. Hubblo. <https://github.com/hubblo-org/scaphandre>. Accessed: 10 July 2024
11. Meross IoT. <https://pypi.org/project/meross-iot/>. Accessed: 10 July 2024
12. Amanusk: s-tui. <https://github.com/amanusk/s-tui>. Accessed: 10 July 2024
13. PyTorch MNIST Examples. <https://github.com/pytorch/examples/tree/main/mnist>. Accessed: 10 July 2024
14. PyTorch Siamese Network Examples. [https://github.com/pytorch/examples/tree/main/siamese\\_network](https://github.com/pytorch/examples/tree/main/siamese_network). Accessed: 10 July 2024

# CO<sub>2</sub> Emissions of AI Applications: An Investigation on its Measurement



Pankhuri Verma, V. Dinesh Reddy, and Marco Aiello

**Abstract** The rapid expansion of Artificial Intelligence (AI) has led to a significant increase in the use of Data Centres (DCs), which are essential for processing and storing vast amounts of data. However, this surge in AI deployment has raised environmental concerns about increased Carbon Dioxide (CO<sub>2</sub>) emissions. Various solutions have been proposed to address the energy efficiency of DCs such as advanced cooling systems or selecting training locations with lower cooling needs or greener power supplies. To achieve further improvements, one needs to be able to measure actual emissions at the code level so that an optimization strategy can be designed and evaluated. To address the issue, we explore an innovative approach to precisely measure the CO<sub>2</sub> emissions of AI applications. By introducing a linear regression energy estimation model based on Performance Monitoring Counters (PMCs) we calculate the CO<sub>2</sub> emission of AI applications. PMCs such as the total number of instructions and the total number of cycles of the computer processor are considered ideal for energy estimation due to their strong correlation with the processor's energy consumption and minimal overhead on resource utilisation. For this research, only the Central Processing Unit (CPU) and Dynamic Random Access Memory (DRAM) are considered, as they consume the maximum energy compared to other parts of the processor. This approach is easily extendable to GPUs. In the presented evaluation, the energy estimation model produced an error of only 0.158% for CPU and 0.272% for DRAM.

**Keywords** Data centres · Artificial Intelligence · Energy consumption · Performance monitoring counters

---

P. Verma · V. Dinesh Reddy · M. Aiello  
Service Computing, IAAS, University of Stuttgart, Stuttgart, Germany  
e-mail: [st180247@stud.uni-stuttgart.de](mailto:st180247@stud.uni-stuttgart.de)

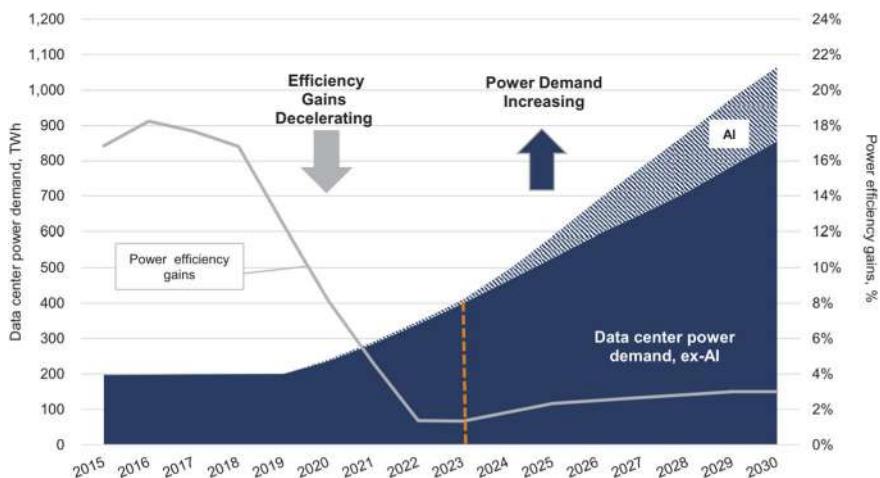
M. Aiello  
e-mail: [marco.aiello@iaas.uni-stuttgart.de](mailto:marco.aiello@iaas.uni-stuttgart.de)

V. Dinesh Reddy (✉)  
CSE, SRM University, Amaravati, AP, Germany  
e-mail: [dinesh.vemula@iaas.uni-stuttgart.de](mailto:dinesh.vemula@iaas.uni-stuttgart.de)

## 1 Introduction

The current success of the AI is calling for substantial expansion in high-performance computing clusters and DCs because of their intensive processing requirements and energy consumption. Advancements in AI models have significantly improved in domains such as machine translation, speech recognition, and object detection. However, AI models are computationally demanding due to their large datasets, extensive model sizes, and numerous parameters used for training. Developing these models also requires thorough experimentation with various hyperparameters, resulting in increased load on the DC processors and consequent CO<sub>2</sub> emissions.

In the analysis presented by Goldman Sachs, the DC power demand has increased from 1–2% in 2022 to 3–4% in 2023 [1]. It is anticipated that the demand for power in DCs will increase by 160% between the years 2023 and 2030. Figure 1 shows an illustration of the DC power demand in terawatt-hours and % gains. The study also contains a prediction about the growth in power demand in DCs resulting in an increase in CO<sub>2</sub> emissions from DCs by more than 100% by 2030 compared to 2022. It is alarming to see the forecast presented by Goldman Sachs stating that power demand from AI will increase by approximately 200 TWh from 2024 to 2030, with AI expected to account for around 20% of the total DC power demand by 2030 [2]. The study also estimates that a ChatGPT search consumes about 6–10 times more power than a traditional Google search. To put things into perspective, the carbon footprint of training LLMs like Bidirectional Encoder Representations from Transformers (BERT) on Graphics Processing Unit (GPU) is comparable to the emissions from a New York to San Francisco flight [3]. As the usage of these models expands, the environmental impact intensifies, contributing to global climate



**Fig. 1** Power demand of DCs over the years [1]

change. Therefore, it is imperative to develop and implement strategies to reduce these emissions.

Historically, developers have focused on energy-efficient scheduling and resource allocation to improve energy utilisation of DCs rather than employing optimisation strategies at the code level [4]. They have primarily emphasised the software qualities such as performance and accuracy of AI applications without considering energy efficiency. This has often led to highly energy-intensive applications. Therefore, it is essential to be aware of energy consumption and to optimise the code. As a step in this direction, the goal and contributions of the present paper are:

- Identifying correlations between PMCs and AI energy consumption.
- Development of a linear regression energy estimation model for AI applications based on PMCs.
- CO<sub>2</sub> estimation of AI applications execution.

The rest of the paper is structured as follows. In Sect. 2, we overview related work about software energy usage. Section 3 contains a discussion about the methodology for using PMCs to determine the energy consumption of AI applications. Section 4 illustrates the methodology regarding CO<sub>2</sub> emission estimation. The paper closes with Sect. 5 summarising the findings of the presented research and discussing directions of future work.

## 2 Related Work

The pioneering work by Tiwari et al. was the first of its kind to use an instruction-based power analysis of software operations [5]. The methodology was groundbreaking because it shifted the focus from traditional hardware-centric power analysis to a software-oriented perspective. These models provided insights into how different instructions impact overall power usage. The research by Bellosa introduced PMCs as an effective indicator of power usage [6]. García-Martín et al. provide an overview of the methods for estimating energy usage of Machine Learning (ML) applications [7]. The research showcased the most recent energy estimation software tools, various methods for estimating energy usage, and energy estimation models. In another paper by García-Martín et al. , the energy estimation techniques were categorised based on the different ML scenarios, like processing big datasets, training, and inference stages. They discussed the various types of analytical and empirical methods used to measure energy consumption. Simulation and PMC were also discussed as the best practices for measuring energy consumption [8].

Contreras and Martonosi proposed a power estimation model for the Intel PXA255 processor that used PMCs to estimate CPU and DRAM power consumption [9]. It linked PMCs such as instructions executed, data dependencies, instruction cache misses, and Translation Lookaside Buffer (TLB) misses with an error rate of 4%. This study attempted to estimate the energy consumption of software by running various

traditional benchmarks to generate energy models. However, AI applications differ from traditional benchmark programs in terms of computational demand and energy usage, due to the large dataset size, various model parameters and weights. Building upon this foundational research, our study focuses specifically on AI applications. Therefore, specific AI models have been executed to gather the dataset for our energy model creation. Unlike other approaches, our work is the first of its kind to use AI models to estimate energy consumption during model training. This approach allows a more precise and tailored understanding of the energy requirements unique to AI workloads.

### 3 PMC-Based Energy Estimation Model

We use PMCs to measure the energy consumption of AI applications. PMCs are dedicated hardware registers used for tracking executed instructions, cache hits, and misses in modern CPUs that monitor various performance-related events. They are used to collect valuable information regarding software and hardware performance attributes. The present research aims to find the correlation between hardware PMCs and the energy consumption of AI applications. Our experiments have been performed on Intel® CoreTM i7-8565U CPU running at 1.80 GHz (142, 0 × 8e). For our analysis, we focus primarily on the training process of AI applications, as it is the most resource-intensive part of model development [10]. Energy consumption of processor components such as the CPU and DRAM is only taken into consideration as they have the most immediate influence on the AI training process. We do not consider components like Solid State Drives (SSDs) and Hard Disk Drive (HDD) as they do not directly influence the running AI processes [9].

#### 3.1 Dataset Generation

Our methodology utilises the process of running various AI benchmark programmes to generate a dataset comprising PMC data and corresponding CPU and DRAM energy measurements. The ML models used in this study to generate our dataset are Linear Regression, Logistic Regression, K-Nearest Neighbour (KNN), Support Vector Machines (SVMs), Decision Trees, and Neural Networks. The dataset comprises values of various PMCs such as total instructions, total cycles, cache hits and misses, and energy consumption of CPU and DRAM. We employ these models to obtain the energy and performance of only the model training stage. This facility is not provided by traditional ML benchmarks currently. Each model was executed with a range of dataset sizes, hyperparameters, and features to represent the various operational profiles of ML applications, and a dataset comprising approximately 2.5k data points was gathered.

We used the widely-known Performance Application Programming Interface (PAPI) interface and the Running Average Power Limit (RAPL) interface to measure the PMC data and energy consumption of CPU and DRAM, respectively.

**PAPI:** The PAPI interface is a portable way to access hardware PMCs. It tracks over 100 predefined events through high-level and low-level interfaces [11]. Our research utilises the PAPI high-level events called PAPI\_TOT\_INS (total instructions) and PAPI\_TOT\_CYC (total CPU cycles) that measure the impact of system modifications on performance and have low overhead due to parallel processing [8].

**RAPL:** The RAPL interface is a feature found in modern Intel processors that monitors power usage of the computing unit such as CPU socket package, DRAM, and GPU [12]. It has been designed to measure the energy consumption of specific code snippets, making it ideal for fine-grained analysis. The primary advantage of RAPL is that it measures energy consumption without interfering with already running computational processes [13].

These Application Programming Interfaces (APIs) were called using methods in the Python package called pyRAPL [14, 15] and pyPAPI to simultaneously measure the energy consumption and PMCs, respectively.

### 3.2 Selection of PMCs

The Intel CoreTM i7-8565U CPU running at 1.80 GHz (142, 0 × 8e) processor provides access to 59 different PMCs via the PAPI interface. However, during our research, we discovered that not all PMCs show a strong correlation with the energy consumption of CPU and DRAM energy. Therefore, a correlation between energy consumption and the PMC data was examined in the dataset using Spearman's Rank Correlation Coefficient ( $\rho$ ). A coefficient of +1 indicates a strong positive correlation, while -1 indicates a strong negative correlation. Based on the  $\rho$  value, we chose only those PMCs that show a strong correlation with energy consumption to avoid redundancy. It was observed that the total CPU cycles and CPU energy have a  $\rho$  of 0.922 and total instructions and CPU energy have a  $\rho$  of 0.861. Similarly, the total CPU cycles and DRAM energy have a  $\rho$  of 0.869 and total instructions and DRAM energy have a  $\rho$  of 0.751. Other PMC events exhibited poor correlation with CPU and DRAM energy, having  $\rho$  values of 0.55 or less. Therefore, we have used only total instructions and total CPU cycles for our energy model creation.

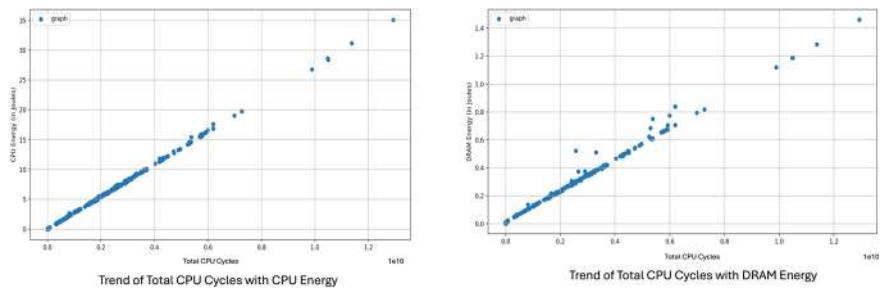
**Total Instructions:** Total instructions are the individual operations carried out by a CPU according to the program.

**Total CPU Cycles:** The number of clock cycles that the CPU uses to complete tasks is measured by CPU cycles.

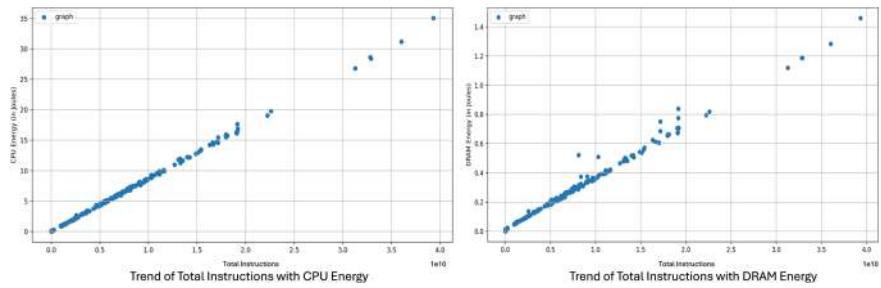
Table 1 shows the dataset comprising total instructions, total CPU cycles, CPU energy, and DRAM energy of 10 data instances.

**Table 1** 10 Data instances of the dataset consisting of total instructions, total CPU cycles, CPU energy, and DRAM energy used to train the linear regression energy estimation model

Index	Total instructions	Total CPU cycles	CPU energy	DRAM energy
1	5,744,069	5,457,638	0.1201597	0.0149658
2	5,704,238	5,362,627	0.0356077	0.0042114
3	5,723,654	5,562,773	0.0313781	0.0048767
4	5,751,021	5,371,274	0.026184	0.0041076
5	5,775,481	6,344,354	0.0481506	0.004895
6	5,811,632	6,053,471	0.0405944	0.004718
7	5,833,511	5,218,797	0.0323302	0.0042725
8	5,858,551	5,349,801	0.0264892	0.0043945
9	5,877,574	5,374,618	0.0291869	0.0042053
10	5,908,991	5,284,704	0.0176758	0.0038942



**Fig. 2** Linear correlation between total CPU cycles and CPU energy and DRAM energy



**Fig. 3** Linear correlation between total instructions and CPU energy and DRAM energy

Supporting the above correlation values, Fig. 2 depicts a perfect linear relationship of CPU cycles with CPU energy and DRAM energy. Similarly, Fig. 3 illustrates a linear relationship of total instructions with CPU energy and DRAM energy. This linear trend aids our argument for the creation of a linear energy model based on total instructions and total CPU cycles.

**Table 2** Values of CPU and DRAM regression model weights

Weight	Value
$b_{1,0}$	0.00042871
$b_{1,1}$	0.84030965
$b_{1,2}$	0.16071597
$b_{2,0}$	0.00153388
$b_{2,1}$	0.84543874
$b_{2,2}$	0.18049153

### 3.3 Linear Regression Energy Estimation Model

We create a linear regression model to estimate the energy usage of AI applications operating on Intel processors. This process involves using the generated dataset to create our linear energy model and determining the model weights. The dataset is passed through the data preprocessing stage, where the extreme outliers are removed and the dataset is normalised using Min-Max Scaling [16] to bring all the features on a uniform, unit scale [0, 1]. Consequently, the dataset is divided into training (55%), validation (25%), and testing (20%) sets. A linear regression algorithm is employed to generate the energy models using total instructions and total CPU cycles as the independent variables and CPU energy and DRAM energy as the dependent variables. The model is fine-tuned using the validation dataset to achieve the highest accuracy and test data is utilised to test the model's accuracy.

The following equations represent the linear model used to estimate the CPU and DRAM energy consumption in micro-joules ( $\mu\text{J}$ ):

$$\text{CPU energy} = b_{1,0} + (b_{1,1} \times \text{tot}_{\text{ins}}) + (b_{1,2} \times \text{tot}_{\text{cyc}}) \quad (1)$$

$$\text{DRAM energy} = b_{2,0} + (b_{2,1} \times \text{tot}_{\text{ins}}) + (b_{2,2} \times \text{tot}_{\text{cyc}}) \quad (2)$$

where

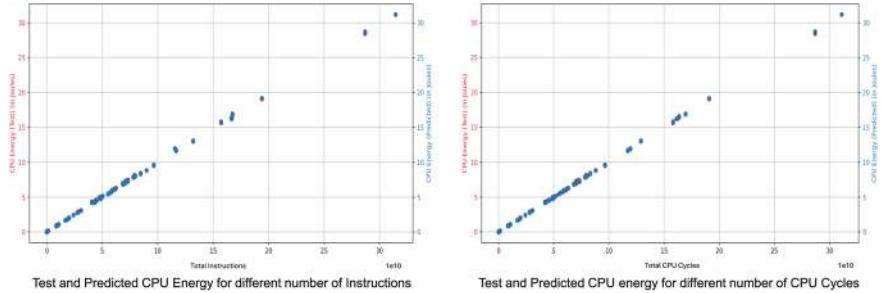
- $\text{tot}_{\text{ins}}$  denotes the total number of instructions executed.
- $\text{tot}_{\text{cyc}}$  denotes the total number of CPU cycles executed.
- $b_{1,0}$ ,  $b_{1,1}$ , and  $b_{1,2}$  represent the regression weights of the CPU model.
- $b_{2,0}$ ,  $b_{2,1}$ , and  $b_{2,2}$  represent the regression weights of the DRAM model.

Table 2 shows the values of CPU and DRAM model weights that will be used to precisely estimate the energy consumption of the CPU and DRAM when an AI model is running.

Using Eqs. 1 and 2 we can estimate the precise value of energy consumption during the training phase of any AI application based on the total number of instructions and total number of CPU cycles.

**Table 3** Metrics for performance evaluation of CPU and DRAM energy models

Energy model	MAE	$R^2$ score	Error (%)
CPU	0.00060	0.9998	0.158
DRAM	0.00255	0.9925	0.273

**Fig. 4** Performance of CPU energy model—predicted versus test energy values

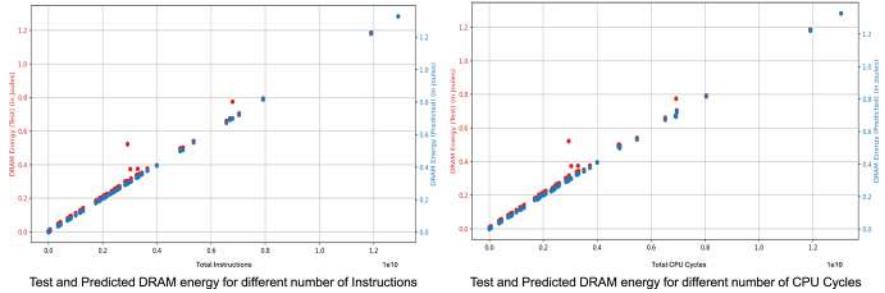
### 3.4 Analysis of Energy Models

The results of our energy models are shown in Table 3. The energy models showcase an error of only 0.158 and 0.273% for CPU and DRAM, respectively. The achieved level of accuracy is notably superior compared to previous research in this domain [9]. Both energy models perform very well with a CPU Mean Absolute Error (MAE) of 0.00060 and DRAM MAE of 0.00255. The R Squared ( $R^2$ ) scores of 0.9998 and 0.9925 for CPU and DRAM, respectively, also support the argument that these models are a good fit for measuring the energy consumption of AI applications.

The graphs in Fig. 4 show the CPU energy model’s performance based on the total number of instructions and total CPU cycles. The blue data points represent the predicted CPU energy while the red data points represent the test CPU energy. A close match between the predicted and test data points validates our model accuracy with only 0.158% error rate. Similarly, the graphs in Fig. 5 show the DRAM energy model’s performance based on the total number of instructions and the total CPU cycles. The predicted (blue) and test (red) DRAM energy data points coincide, with an error rate of only 0.273%.

## 4 CO<sub>2</sub> Emission Estimation of AI Applications

The next step is the assessment of CO<sub>2</sub> emissions resulting from the AI applications. There is a non-strict relationship between CO<sub>2</sub> emissions and energy usage due to various factors such as geographic location, and energy mix [17]. We use the



**Fig. 5** Performance of DRAM energy model—predicted versus test energy values

carbon intensity in our analysis to accurately represent the geographical variations in emissions.

**Carbon Intensity** is a coefficient showing the weight of CO<sub>2</sub> emissions, expressed in kilogramme (kg), for each Kilowatt-hour (KWh) of electricity produced. The carbon intensity is determined by the energy mix of a region that includes fossil fuels and renewable energy sources like solar power, biomass, and more [17].

The carbon intensity of a geographical region can be used to measure the precise CO<sub>2</sub> emissions. Equation 3 captures the way to precisely calculate the CO<sub>2</sub> emission of a computing unit by multiplying the total CPU and DRAM energy consumption by, for instance, Germany's carbon intensity.

$$\text{Total CO}_2 \text{ Emission} = \sum (\text{Energy}_i) \times \text{Carbon Intensity} \quad (3)$$

where

- $i$  represents the computing unit (CPU and DRAM).
- Energy<sub>i</sub> denotes the energy consumed by the computing unit (in KWh).
- Carbon Intensity denotes the carbon intensity of the corresponding region.

Since our experiments are conducted on an Intel machine running at our university in Germany, the carbon intensity of Germany was used to measure the precise value of CO<sub>2</sub> emissions from any AI application. As of April 29, 2024, the carbon intensity coefficient for the German region is 385.389, according to Codecarbon [18].

To illustrate our methodology of calculating CO<sub>2</sub> emission, we have run a simple linear regression model on the sklearn California housing dataset. The total instructions executed while running the model are 1884248.8 and the total CPU cycles are 2219907.

We use our generated formula for CPU and DRAM energy to calculate the total energy consumption and CO<sub>2</sub> emission of the model.

$$\begin{aligned}
\text{CPU energy} &= b_{1,0} + (b_{1,1} \times \text{tot}_{\text{ins}}) + (b_{1,2} \times \text{tot}_{\text{cyc}}) \\
&= 0.00042871 + (0.84030965 \times 1884248.8) \\
&\quad + (0.16071597 \times 2219907) \\
&= 1939719.34 \mu\text{J} \\
&= 1.93971934 \text{ J}
\end{aligned} \tag{4}$$

$$\begin{aligned}
\text{DRAM energy} &= b_{2,0} + (b_{2,1} \times \text{tot}_{\text{ins}}) + (b_{2,2} \times \text{tot}_{\text{cyc}}) \\
&= 0.00153388 + (0.84543874 \times 1884248.8) \\
&\quad + (0.18049153 \times 2219907) \\
&= 1993897.47 \mu\text{J} \\
&= 1.99389747 \text{ J}
\end{aligned} \tag{5}$$

$$\begin{aligned}
\text{Total CO}_2 \text{ Emission} &= \sum (\text{Energy}_i) \times \text{Carbon Intensity} \\
&= ((1.93971934 + 1.99389747)/3600000) \times 385.389 \\
&= 0.000421103 \text{ kg}
\end{aligned} \tag{6}$$

Based on Eq. 4, we calculate the energy consumption of CPU, which is 1.93971934 J. The DRAM energy was calculated as 1.99389747 J (Eq. 5). Further, the total energy was used to calculate the CO<sub>2</sub> emission. Energy in Joules is converted to KWh for CO<sub>2</sub> emission calculation. According to Eq. 6, 0.000421103 kg of CO<sub>2</sub> was emitted to train a linear regression model on the California housing dataset.

Our streamlined approach simplifies CO<sub>2</sub> emission estimation, making it valuable for AI development and deployment. This PMC-based energy model sets a new standard in AI energy management for the development of sustainable AI applications. This detailed level of analysis facilitates the reduction of energy consumption, leading to more efficient and eco-friendly AI systems.

## 5 Conclusions

Energy consumption is an important factor to consider when developing ML algorithms. The majority of research focus on improving the accuracy of algorithms while neglecting their energy requirements. With the present research, we aim to give a new perspective on the AI industry by providing insights into the energy consumption pattern of ML models. Our research contributes to the scientific understanding of AI energy consumption and offers practical methodologies that can help developers precisely measure the energy consumption of AI applications based on PMCs.

From our research results, one can conclude that processor-specific PMCs and AI applications' energy consumption are directly correlated. A framework for estimating the energy consumption and, consequently, the CO<sub>2</sub> emissions of different AI models was provided by the use of a linear regression energy model. By applying these regression models, developers can predict and manage the energy consumption of AI operations more effectively, leading to significant energy savings and reducing the environmental impact associated with running intensive AI tasks in the real world.

The research done so far has established an initial understanding of energy consumption and CO<sub>2</sub> emissions in AI applications on CPUs. However, future work will focus on GPUs and TPUs as they have better processing capabilities and are commonly used in the field of AI. Further study will be done on GPU and TPU-based PMCs to estimate the energy consumption of AI applications.

## References

1. Singer B, Bingham DR, Corbett B, Davenport C, Gandol A (2024) AI/data centers' global power surge and the sustainability impact. <https://www.goldmansachs.com/intelligence/pages/gs-research/ai-data-centers-global-power-surge-and-sustainability-impact/report.pdf>. Accessed 21 May 2024
2. Davenport C, Singer B, Mehta N, Lee B, Mackay J. AI, data centers and the coming us power demand surge. <https://www.goldmansachs.com/intelligence/pages/gs-research/generational-growth-ai-data-centers-and-the-coming-us-power-surge/report.pdf>. Accessed 21 May 2024
3. Strubell E, Ganesh A, McCallum A (2019) Energy and policy considerations for deep learning in NLP
4. Dinesh Reddy V, Gangadharan GR, Rao GS VRK, Aiello M (2020) Energy-efficient resource allocation in data centers using a hybrid evolutionary algorithm. 71–92
5. Tiwari V, Malik S, Wolfe A, Lee MT-C (1996). Instruction level power analysis and optimization of software, pp 326–328. <https://doi.org/10.1109/ICVD.1996.489624>
6. Bellosa F (2000) The benefits of event-driven energy accounting in power-sensitive systems
7. García-Martín E, Rodrigues CF, Riley G, Grahn H (2019) Estimation of energy consumption in machine learning. *J Parallel Distrib Comput* 134:75–88
8. García-Martín E, Lavesson N, Grahn H, Casalicchio E, Boeva V (2019) How to measure energy consumption in machine learning algorithms. In: ECML PKDD 2018 workshops, Cham. Springer, pp 243–255
9. Contreras G, Martonosi M (2005) Power prediction for intel xscale ® processors using performance monitoring unit events. In: Proceedings of the 2005 international symposium on low power electronics and design, ISLPED '05, New York, NY, USA. Association for Computing Machinery, pp 221–226
10. Haghshenas K, Setz B, Aiello M (2022) CO<sub>2</sub> emission aware scheduling for deep neural network training workloads. In: 2022 IEEE International conference on big data (big data), pp 1542–1549
11. Phil M, Shirley M, Christine D, Ho G (1999) PAPI: A portable interface to hardware performance counters
12. Mazouz A, Wong DC, Kuck D, Jalby W (2017) An incremental methodology for energy measurement and modeling. In: Proceedings of the 8th ACM/SPEC on international conference on performance engineering, ICPE'17, New York, NY, USA. Association for Computing Machinery, pp 15–26
13. University of Maine System (2022) Running average power limit energy reporting. <https://www.intel.com/content/www/us/en/developer/articles/technical/software-security-guidance/>

- [advisory-guidance/running-average-power-limit-energy-reporting.html](#), 2022. Accessed 22 May 2024
- 14. University of Lille (2019) Welcome to PyRAPL's documentation! <https://pyrapl.readthedocs.io/en/latest/>. Accessed 22 May 2024
  - 15. University of Lille (2018) PyrRAPL version 0.2.3.1 project description page. <https://pypi.org/project/pyRAPL/>. Accessed 22 May 2024
  - 16. de Amorim LBV, Cavalcanti GDC, Cruz RMO (2023) The choice of scaling technique matters for classification performance. *Appl Soft Comput* 133:109924
  - 17. Lottick K, Susai S, Friedler SA, Wilson JP (2019) Energy usage reports: environmental awareness as part of algorithmic accountability
  - 18. CodeCarbon (2020) Germany carbon intensity. [https://github.com/mlco2/codcarbon/blob/master/codcarbon/data/private\\_infra/global\\_energy\\_mix.json](https://github.com/mlco2/codcarbon/blob/master/codcarbon/data/private_infra/global_energy_mix.json). Accessed 28 Apr 2024

# Research on Optimization of Machine Translation Performance Based on Deep Learning Algorithm



Yan Meng and Jiuquan Zhang

**Abstract** This article aims to solve the problem of scarce data from a single corpus and difficulty in learning the correspondence between the source and target languages. It uses a multi-translation parallel corpus for neural machine translation (NMT) research: extracting data from multiple parallel corpora and cleaning them, constructing a NMT model based on the transformer architecture. It uses Xavier for parameter initialization and unfolds model training through backpropagation and stochastic gradient descent algorithms. This article adopts the bagging method to integrate different parameter models and optimizes translation results based on language models and phrase tables. The research results indicate that the Bilateral Evaluation Understudy (BLEU) of the paper's model in the Europarl corpus is 0.84, which is approximately 5.0% higher than the Generative Pre-trained Transformer (GPT). The improvement method adopted can achieve more accurate NMT in data scarcity scenarios.

**Keywords** Neural machine translation · Parallel corpus · Language model-based optimization · Transformer architecture

## 1 Introduction

In the information age, machine translation, as an important branch of language technology, plays a bridging role in connecting different languages and cultures. Traditional NMT models often encounter problems such as data scarcity and difficulty in capturing the correspondence between source and target languages when facing a single corpus, which limits their performance and generalization ability. The existence of problems affects the accuracy and fluency of machine translation systems, while also restricting their practical application scope. Seeking a solution to the current challenges is particularly urgent.

---

Y. Meng · J. Zhang (✉)

School of Foreign Languages, Huainan Normal University, Huainan, China  
e-mail: [8861531@qq.com](mailto:8861531@qq.com)

The research focuses on building efficient and accurate NMT models. Data can be obtained from parallel corpora of multiple translations, covering multiple languages and thematic domains, and processed and cleaned to ensure data quality. This article adopts a multi-layer Transformer Encoder-Decoder structure, utilizes multi-head self-attention mechanism and feedforward neural network to capture semantic information in the input sequence, and trains the model through backpropagation algorithm and random gradient descent algorithm. It combines the bagging method with multiple NMT models with different initialization parameters to improve the overall robustness and accuracy of translation. This article uses sentence-level rearrangement based on language models and word replacement based on phrase lists to post-process translation results and further optimize translation quality. The research content innovatively improves the performance and stability of NMT models, providing more reliable support for cross language communication.

## 2 Related Works

In existing research, many scholars have realized the limitations of a single corpus on machine translation performance and have attempted to solve this problem by introducing multimodal data, enhancing learning, and other methods. By utilizing object interactions in space and time to eliminate ambiguity, Li et al. [1] successfully improved and validated the Unsupervised Machine Translation (UMT) model. Regarding the translation delay caused by autoregressive mechanisms, researchers such as Chenze [2] applied sequence level objectives to train the NMT model. In the field of multimodal sentiment analysis, Fan et al. [3] proposed the Transformer-Based Encoding-Decoding Translation Network (TEDT) model with an accuracy of 89.3%. However, these attempts often face challenges such as high data acquisition costs, difficulty in annotation, and poor performance in practical applications [4, 5]. At present, there are still many problems in how to more efficiently utilize corpus data for NMT research.

The use of multi-translation parallel corpora for NMT can effectively improve model performance, especially in situations where data is scarce. Scholars such as Hyeonseok [6] have experimentally verified that the Multilingual Bidirectional and Autoregressive Transformers (mBART) model can achieve excellent machine translation results with only a small amount of data when filtering through parallel corpora. He considered the balance between parallel corpora in machine translation, and scholars such as Chanjun [7] proposed a new Corpus Weight Balance (CWB) method, which can construct high-quality parallel corpora using monolingual corpora. Based on Gated Recurrent Unit (GRU) and Transformer network, researchers such as Sukanta [8] proposed an improved NMT system for use in resource scarce situations. However, existing methods often overlook translation differences and language styles between different versions, leading to a decrease in the model's generalization ability [9, 10]. The study can explore the efficient

application of multi-translation parallel corpora to improve the performance and generalization ability of NMT.

### 3 Implementation of NMT

#### 3.1 *Data Collection and Cleaning*

The database extracts data from many well-known resources. The site covers many different languages, different disciplines, and is used to build a diverse, colorful model. The TED Dialogue includes public presentations on a variety of topics, Europarl includes presentations and discussions at several European Parliament congresses, and the UN Bilingual Corpus contains UN and official documents in multiple languages. On this basis, this paper proposes a method based on sentence length filtering, which ensures the consistency and reasonableness of training samples and prevents the influence of too long or too short sentences. In the process of de-duplication, the hash table is used to make the data in the de-duplication process which have diversity and independence. On this basis, a speech recognition algorithm based on linguistic model is used to eliminate the noise, thus effectively improving the data purity and quality. Aiming at the problem of large sample size among different languages, this paper proposes a resampling and weighting method to achieve the relative balance of sample size among different languages.

#### 3.2 *Build a NMT Model*

The study adopts a multi-layer Transformer Encoder-Decoder structure. The encoder section utilizes multi-head self-attention mechanism and feedforward neural network, which can effectively capture long-distance dependencies in the input sequence and extract semantic information. The decoder section also uses multi-head self-attention mechanism, attention mechanism, and feedforward neural network to decode the encoder output and generate the target language sequence. He adopts the Xavier initialization method, which automatically adjusts the initial weight values based on the number of input and output neurons, keeping the input and output of the network at the same level. This reduces the problem of vanishing or exploding gradients during the training process, accelerates convergence, and improves model stability and performance. Through the backpropagation algorithm, the model updates its parameters based on the gradient information of the loss function, gradually converging to the optimal solution. The stochastic gradient descent algorithm estimates gradients by randomly selecting subsets of training data and updating model parameters in each iteration, accelerating training speed and reducing computational costs.

The multi-head self-attention mechanism is represented by the following formula:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \quad (1)$$

Here,  $Q, K, V$  represent the matrix representation of query, key, and value, respectively.  $d_k$  represents the dimension of the key. Softmax normalization can be performed by calculating the dot product of the query and key and finally multiplied by the value to obtain the output. For the Xavier initialization method, it is represented as:

$$W_{ij} \sim U\left[-\frac{\sqrt{6}}{\sqrt{n_i + n_j}}, \frac{\sqrt{6}}{\sqrt{n_i + n_j}}\right]. \quad (2)$$

Among them,  $W_{ij}$  is the element in the  $i$  row and  $j$  column of the weight matrix, and  $n_i$  and  $n_j$  are the number of input and output neurons in the weight matrix, respectively.

The backpropagation algorithm involves the calculation of gradients and parameter updates, which are represented by the chain rule: assuming that the loss function is  $L$  and the parameter is  $\theta$ ; the parameter update formula is as follows:

$$\theta_{t+1} = \theta_t - \alpha \nabla_{\theta_t} L. \quad (3)$$

Here,  $\alpha$  is the learning rate, and  $\nabla_{\theta_t} L$  is the gradient of the loss function relative to the parameter. In the stochastic gradient descent algorithm, the formula for updating parameters in each iteration is:

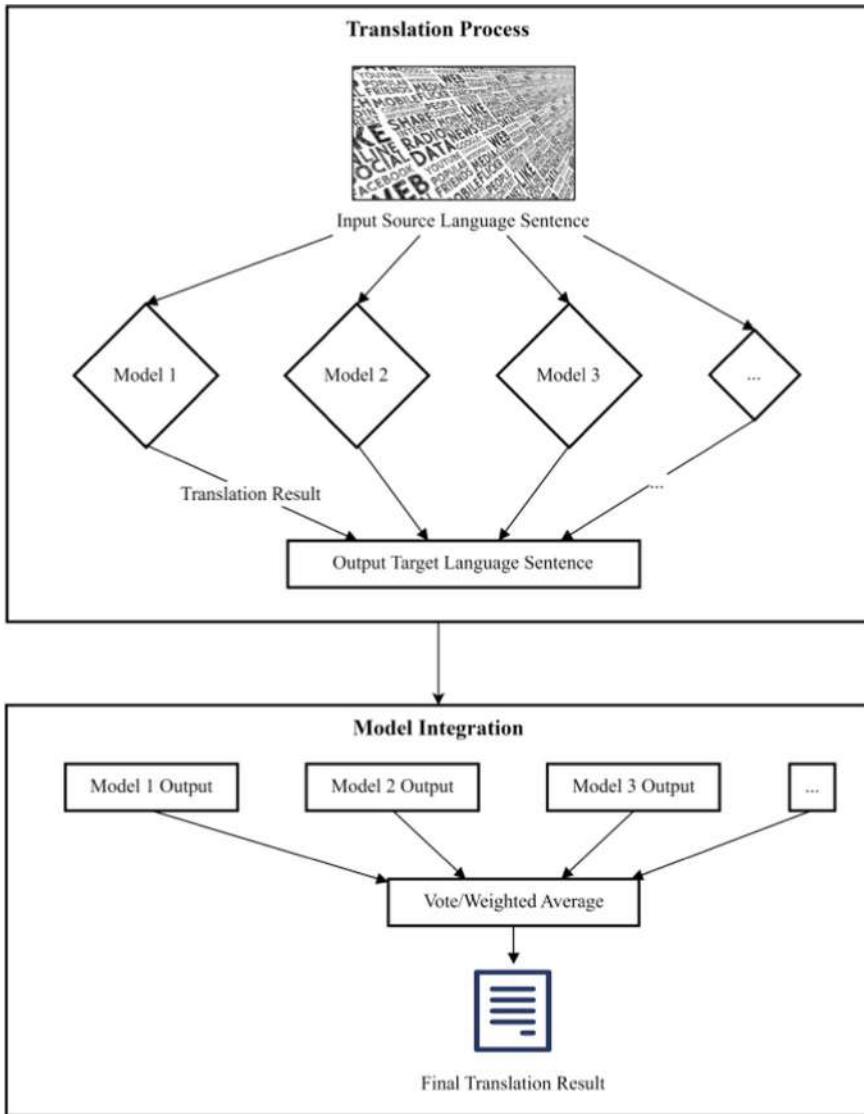
$$\theta_{t+1} = \theta_t - \alpha \nabla_{\theta_t} L(x_{i:t}, y_{i:t}). \quad (4)$$

Among them,  $x_{i:t}$  and  $y_{i:t}$  represent subsets of the input and target sequences, respectively, and  $\nabla_{\theta_t} L(x_{i:t}, y_{i:t})$  is the gradient of the loss function on the subset.

### 3.3 Multi-model Integration

As shown in Fig. 1, a NMT model is constructed using the Bagging method combined with multiple different initialization parameters. By training multiple independent NMT models, each with different initialization parameters, their outputs are voted/weighted average to improve the overall robustness and accuracy of translation.

In the study, multiple independent NMT models were trained. The model adopts the same architecture, but uses different initialization parameters to ensure their differences. During the training process, each model learns different features and patterns on the training data, forming diversity. The output of independent models



**Fig. 1** Multi-model integration process

can be summarized. In the voting method, each model translates the input source language sentences and votes based on its output, selecting the translation with the most votes as the final result. In the weighted average method, weights can be assigned based on the performance of each model on the validation data, and the output can be weighted average according to the weights to obtain the final translation result. Model integration can effectively reduce the error of single model and improve the quality

and accuracy of overall translation. In the face of complex language phenomena and data scarcity, model integration can play a greater advantage. By combining multiple models with different initialization parameters, the diversity of the models can be further increased, and the translation effect can be improved, making the models more robust and reliable.

### 3.4 Post Processing of Translation Results

Language model-based rearrangement methods can be used to rearrange translation results at the sentence level. Assuming that the original translation result is  $T$  and the language model of the target language is LM, the goal is to maximize the probability of sentences in the target language. The optimization objective is expressed as:

$$\arg \max_T P(T|LM). \quad (5)$$

Among them,  $P(T|LM)$  represents the probability of the translation result  $T$  given the target language model. Use the Viterbi algorithm (VA) to find the sentence rearrangement sequence that maximizes the probability. Adopting a phrase table based word replacement method to further optimize translation results. Set the original translation result as  $T$ , with the goal of minimizing translation quality issues caused by inaccurate word selection in the translation result. Express the optimization problem as:

$$\arg \min_{T'} \sum_{i=1}^{|T|} \text{penalty}(T_i, T'_i). \quad (6)$$

Here,  $T'$  represents the translation result after word replacement, and  $T_i$  and  $T'_i$  represent the  $i$  th word in the original translation result and the replaced translation result, respectively. The penalty function represents the penalty caused by inaccurate word selection. By constructing a penalty matrix, the degree of inaccurate word selection can be quantified, and words in the translation results can be replaced to minimize the penalty value based on the phrase correspondence in the phrase table. This article uses heuristic search algorithms to solve the above optimization problems and achieve word replacement in translation results.

**Table 1** Comparison of various models under BLEU indicators

Model	TED talks	Europarl	UN Parallel Corpus
RNN	0.65	0.68	0.62
CNN	0.72	0.70	0.68
RNN	0.68	0.67	0.64
Bahdanau	0.75	0.72	0.70
BERT	0.80	0.78	0.75
GPT	0.82	0.80	0.78
mBERT	0.85	0.82	0.80
This paper	0.87	0.84	0.82

## 4 Results and Discussion

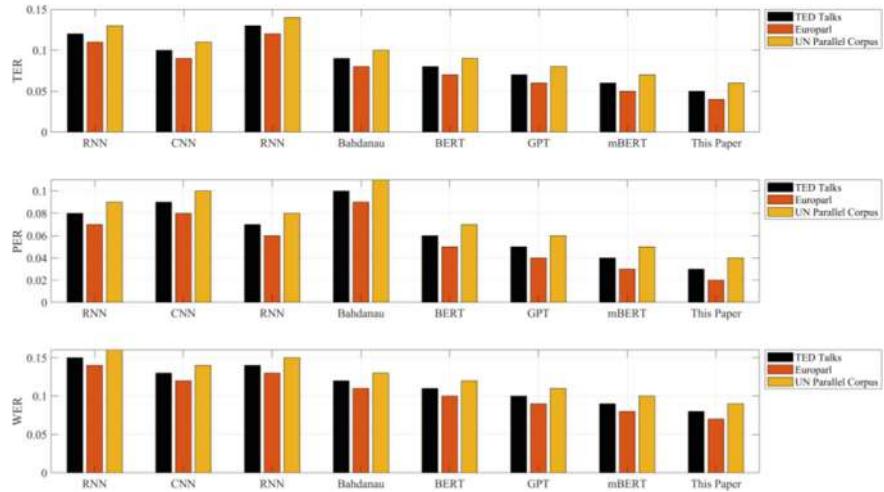
### 4.1 BLEU Evaluation

BLEU indicators can be used to quantify the quality of translation results. To measure translation accuracy, the degree of matching between machine translation results and human reference translation is compared. The machine translation results can be compared with the reference translation to calculate the BLEU score for each sentence, and the average is taken as the overall BLEU score for the translation. This article evaluates the performance of NMT systems on different corpora (TED Talks, Europarl, UN Parallel Corpus) and compares the paper's model with Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Recursive Neural Network (RNN), Bahdanau attention-based translation model, Bidirectional Encoder Representations from Transformers (BERT), GPT, and Multilingual Bidirectional Encoder Representations from Transformers (mBERT). The results of comparing their representations from transformers' model are shown in Table 1.

Compared to the GPT and mBERT models that perform better in other models, this model achieves higher translation quality. In the TED Talks corpus, BLEU reached 0.87, which is about 2.4% higher than mBERT and about 6.1% higher than GPT. In the Europarl corpus, BLEU reaches 0.84, which is about 2.4% higher than mBERT and about 5.0% higher than GPT. In the UN Parallel Corpus, BLEU reaches 0.82, which is approximately 2.5% higher than mBERT and 5.1% higher than GPT.

### 4.2 TER, PER, WER Evaluation

This article uses indicators such as Translation Edit Rate (TER), Position-independent Edit Rate (PER), and Word Error Rate (WER) to evaluate translation



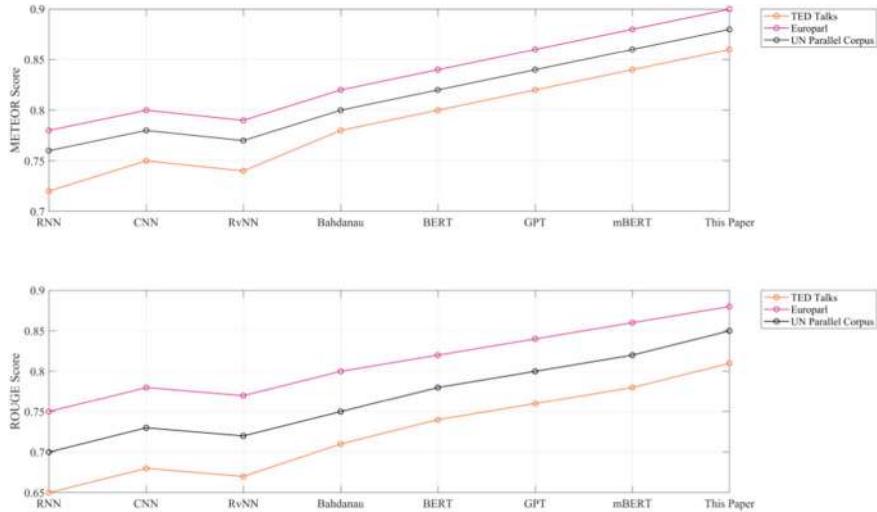
**Fig. 2** Evaluation results of various translation models

results, measuring the editing distance between machine translation results and reference translations, reflecting the fluency and accuracy of translation. The result is shown in Fig. 2.

Figure 2 provides performance evaluation results of different NMT models under three different evaluation metrics in three different corpora. The model in this article performs best in all corpora and evaluation metrics, achieving good results in all aspects, including the number of editing operations, the number of position-independent editing operations, and the error rate at the word level. In the TED Talks corpus, the TER of the paper's model is 0.05, PER is 0.03, and WER is 0.08, which is significantly better than other models. Under the Europarl and UN Parallel Corpus corpora, the model also showed the best performance. At the same time, it is noted that BERT, GPT, and mBERT models also perform well in various corpora, but there is still a certain gap compared to the model proposed in this paper. The translation model introduced in this article performs well in multiple evaluation metrics and different corpora, surpassing other common NMT models. This validates the effectiveness of the research method and improvement methods in this article, providing a powerful solution for solving the translation problem of data scarcity.

### 4.3 METEOR and ROUGE Evaluation

In this paper, evaluation indicators such as Metric for Evaluation of Translation with Explicit Ordering (METEOR) and Recall-Oriented Understanding for Gisting Evaluation (ROUGE) are further used to comprehensively evaluate the translation results. The METEOR indicator mainly considers the degree of WORD matching



**Fig. 3** Evaluation results of each model under METEOR and ROUGE indicators

and SENTENCE FLUENCY between THE translation result and the reference translation, while the ROUGE indicator is mainly used to EVALUATE THE QUALITY OF THE TEXT summary. The result is shown in Fig. 3.

The upper and lower subgraphs in Fig. 3 represent the evaluation results of the model for the METEOR and ROUGE indicators, respectively. In the TED Talks corpus, the model with the highest METEOR score in this article reached 0.86, while the RNN model had the lowest score of 0.72. The model with the highest ROUGE score is also the one in this article, with a score of 0.81, while the RNN model with the lowest score is 0.65. Under the Europarl corpus, the model proposed in this article performed the best again, with a METEOR score of 0.90 and a ROUGE score of 0.88. Under the UN Parallel Corpus: Similar to the first two corpora, the model proposed in this paper performs best on METEOR and ROUGE, with scores of 0.88 and 0.85, respectively. Regardless of the corpus, the translation model proposed in this article performs best on the comprehensive evaluation metrics METEOR and ROUGE, while the traditional RNN model performs relatively poorly. The data further confirm the superiority of the proposed model in different corpora.

## 5 Conclusions

The study explored the issues of NMT models in single corpus data scarcity and difficulty in learning the correspondence between source and target languages. By using a multi-translation parallel corpus for research, data were extracted and cleaned from multiple sources, and an efficient and accurate NMT model was constructed.

The experimental results have demonstrated that the improved method adopted has high accuracy and performance in data scarcity scenarios. However, the model may experience performance degradation when dealing with translation in specific fields or languages, and further targeted improvements are needed. Future research can further explore the application of multi-translation parallel corpora, improve models to enhance generalization ability, and promote the wider and deeper development of NMT in practical applications.

**Acknowledgements** The 2022 Anhui Province Education and Teaching Reform Research Key Project “Research on the Cultivation and Practice of Teaching Wisdom for English Normal Students Based on OBE Concept” (2022jyxm1430); The 2022 Huainan Normal University University Level Quality Engineering Project Ideological and Political Construction and Research of College English Curriculum under the Background of New Liberal Arts (2022hsjxm24). “Research on the Translation and Introduction of Tofu Culture from the Perspective of Translation Theory in Anhui Province’s Key Project of Philosophy and Social Sciences in 2023” (2023AH051525) and “Research and Practice on Strategies for Reducing Burden and Improving Quality in Middle School English Teaching under the Background of” Double Reduction “in University Level Humanities and Social Sciences in 2022” (2022XJYB053).

## References

- Li M, Huang P-Y, Chang X, Hu J, Yang Y, Hauptmann A (2022) Video pivoting unsupervised multi-modal machine translation [J]. IEEE Trans Pattern Anal Mach Intell 45(3):3918–3932
- Chenze S, Yang F, Jinchao Z, Fandong M, Jie Z (2021) Sequence-level training for non-autoregressive neural machine translation [J]. Comput Linguist 47(4):891–925
- Fan W, Tian S, Yu L, Liu J, Wang J, Li K et al (2023) TEDT: transformer-based encoding-decoding translation network for multimodal sentiment analysis[J]. Cogn Comput 15(1):289–303
- Meetei LS, Singh TD, Bandyopadhyay S (2024) Exploiting multiple correlated modalities can enhance low-resource machine translation quality [J]. Multimedia Tools Appl 83(5):13137–13157
- Ronan K, Heng H (2022) Empowering learners of English as an additional language: translanguaging with machine translation[J]. Lang Educ 36(6):544–559
- Hyeonseok M, Chanjun P, Sugyeong E, Jeongbae P, Heuiseok L (2021) Filter-mBART based neural machine translation using parallel corpus filtering[J]. J Korea Convergence Soc 12(5):1–7
- Chanjun P, Kinam P, Hyeonseok M, Sugyeong E, Heuiseok L (2021) A study on performance improvement considering the balance between corpus in neural machine translation[J]. J Korea Convergence Soc 12(5):23–29
- Sukanta S, Mohammed H, Asif E, Pushpak B, Andy W (2021) Neural machine translation of low-resource languages using SMT phrase pair injection[J]. Nat Lang Eng 27(3):271–292
- Rejwanul H, Chao-Hong L, Andy W (2021) Recent advances of low-resource neural machine translation[J]. Mach Transl 35(4):451–474
- Ranathunga S, Lee E-SA, Prifti Skenduli M, Shekhar R, Alam M, Kaur R (2023) Neural machine translation for low-resource languages: a survey[J]. ACM Comput Surv 55(11):1–37

# Using Big Data Analytics and Business Intelligence for Flight Delay Prediction



Mona Hassan Asiri, Abdullah S. A. L.-Malaise AL-Ghamdi,  
Ayman G. Fayoumi, and Mahmoud Ragab

**Abstract** Flight delays are the airline sector's most serious concern because they disrupt airlines, passengers, and airports. This study examines the use of big data analytics with machine learning and business intelligence to improve flight delay prediction accuracy. The aim is to analyze and visualize flight delays using big data applied to all arriving and departing domestic flights from January 2019 to August 2023 in the United States by using big data analytics and business intelligence. In addition, it aims to predict flight delays and compare six classification algorithms using big data analytics. This analysis and visualization showed that Aircraft Arriving Late and Air Carrier Delay reasons have the highest percentage of delays compared to other reasons, followed by National Aviation System Delay and weather delay. The security delay has the lowest percentage. The year 2023 has the highest number of delays, and June has the highest number of delays. There are many prediction models for flight delays; however, more development of accurate prediction models is still needed. In this study, the selected models were "Random Forest classifier," "Logistic Regression classifier," "Neural Networks classifiers," "Support Vector Machine (classifier)," "Gradient Boosting classifier," and "K-Nearest Neighbors" classifier. The comparison used the resulting confusion matrix to summarize prediction results for the following classification problems: F1-score, accuracy, recall, and precision.

---

M. H. Asiri (✉) · A. G. Fayoumi

Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

e-mail: [mhussainalasiri@stu.kau.edu.sa](mailto:mhussainalasiri@stu.kau.edu.sa)

A. G. Fayoumi

e-mail: [afayoumi@kau.edu.sa](mailto:afayoumi@kau.edu.sa)

A. S. A. L.-M. AL-Ghamdi

Information Systems Department, School of Engineering, Computing and Design, Dar Al-Hekma University, Jeddah, Saudi Arabia

e-mail: [aalmalaise@kau.edu.sa](mailto:aalmalaise@kau.edu.sa)

M. Ragab

Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

e-mail: [mragab@kau.edu.sa](mailto:mragab@kau.edu.sa)

The model performance evaluation results showed that the Support Vector Machine classifier yielded better results than other models, followed by the Neural Networks classifiers, *K*-Nearest Neighbors classifier, Random Forest classifier, and the Logistic Regression classifier showed the worst results.

**Keywords** Big data · Business intelligence · Flight delay · Data visualization · Data analysis · Prediction

## 1 Introduction

“Big data” and “big data analytics” significantly benefit management, business, research, and innovation. “Big data analytics” can give companies an advantage in the market. “Big data” and its related technologies impact business and affect academia and other enterprises that use regular data analytics and business analytics [1]. “Big data analytics” is an emerging technology that has become a leading technology used widely in many fields to facilitate decision-making driven by big data to get the desired business outcomes for individuals and businesses [2].

Since 1989, “business intelligence” (BI) has received significant attention from many fields [3]. BI technology is important for enterprises developing e-services, business, and e-commerce. In addition, enterprise, marketing, organizational, and management intelligence have also used BI technology for development [4]. BI faces challenges due to the fast improvement of “big data” and its technologies. This challenge is how to enhance BI by using “big data analytics” services, which is considered a significant issue for information systems, e-commerce, e-services, and business [2].

Airlines’ most significant concern is flight delays. Flight delays can weaken and harm airlines, passengers, and airports. The “analysis,” “visualization,” and “prediction” of flight delays became a main step in the process of making decisions in the airline sector. The enormous number of flights every day creates a lot of data. Handling and manipulating big data need an effective, fast, and distinctive method for making appropriate decisions at the right time. There are many prediction models for flight delays, and the development of an accurate prediction model is needed. In addition to an accurate prediction model, a distinctive method for dealing with big data for analysis and visualization in an easy, clear, and effective way is also required. This is a complicated process because the air transportation system is complex.

This study aims to analyze, visualize, and predict the flight delays of all scheduled flights from January 2019 to August 2023 in the “United States” by using “big data analytics” and “business intelligence” tools. This integration for flight delay analysis, visualization, and prediction can provide deep analysis and support to increase business growth and improve the airline sector. In this study, the prediction used both the arrival and departure flights.

The objectives of this study are to analyze and visualize flight delays before prediction and compute the correlation between features using big data analytics and

business intelligence tools. In addition, they are to predict flight delays and compare the six classification algorithms using big data analytics. The selected algorithms will be measured using F1-score, accuracy, recall, and precision.

The contribution is the analysis and visualization of flight delays using “big data analytics” and “business intelligence”. In addition, it explores the efficiency of using big data analytics and machine learning to enhance flight delay prediction by comparing the accuracy of six machine-learning models.

The overall layout of this paper is as follows: “Literature review, method, implementation, result and discussion, and conclusion.” The “Literature Review” section is a brief summary of some previous studies that focused on the same domain as this study. The “Method” section provides a detailed description of the case study. The “Implementation” section details the setup and description of the dataset used for implementation. In the “Results and Discussion” section, the implementation results have been discussed. Finally, the conclusion section serves as the final part of this study, summarizing the main findings and offering suggestions for future work.

## 2 Literature Review

Big data analytics is gathering, arranging, and analyzing a large amount of data to explore knowledge [5]. In summary, business objectives and decision-making can be realized by “big data analytics” [6]. Business intelligence refers to a collection of technologies that improve the performance of the enterprise. New challenges and opportunities for business intelligence exist because of the rapid improvement of “big data” and “big data technologies” [6]. For this reason, how to use “big data analytics” to improve BI has become a significant point for information systems. According to the previous discussion, “big data analytics” can be part of BI [7]. A business’s value can be enhanced using “big data analytics” in BI. To enhance the effective use of “big data analytics” to make business decisions, data analysts and decision-makers should collaborate. The decision processes must be accurately managed to reduce possible misunderstandings. Appropriate and efficient analytics methods are needed to leverage the unstructured big data. Utilizing advanced technologies for predictive analytics is essential when dealing with structured big data [5].

There are many applications of “big data analytics” tools that help decision-making. The main big data methods are sentiment analysis, predictive analytics, and social media for active market management, depending on customers, employees, and events, and exploring proactive market influence by intelligence [8]. The most well-known objectives of these applications are predictive, descriptive, and improvement. Big data analytics are currently being used for flight delay predictions. There are many prediction models for flight delays, and the development of accurate models has become difficult because of the difficulty of the air transportation systems and the number of prediction methods. Flight delays are unavoidable and significantly affect

airlines' profitability [9]. Accurately predicting and estimating flight delays are critical for airlines because they can positively impact customer satisfaction and experience and increase airlines' revenue. Numerous studies have been undertaken to build models for forecasting flight delays, with most attempting to predict delays by identifying significant features and characteristics. However, the majority of the proposed approaches have not been accurate enough to handle the large volume of data, dependencies, and extreme number of parameters effectively. Numerous machine-learning algorithms have been suggested for predicting flight delays [10]. The majority of these studies use one of three methods: "Binary classifiers," which predict whether a flight will be delayed; "multi-class classifiers," which predict multiple delay classes; or regression, which estimates the numerical value of the flight delay.

Due to the rapid and significant developments in aviation, the number of departure and arrival flights also increased. In addition, flight delays also increased for several reasons. There are numerous approaches to studying the causal factors of flight delays, and their results and variables are varied. Table 1 shows some studies of flight delay factors using traditional methods.

Several studies used machine-learning algorithms for different objectives related to flight delays and their factors. Table 2 shows some studies using machine algorithms that studied the factors of flight delays.

In [18], a taxonomy of the problems of flight delay prediction is proposed and summarized, and initiatives to address the issues are presented. The authors observed two main categories for forecasting flight delay: Root delay and cancellation and delay propagation. The main methods for flight delay prediction are divided into five categories and have been grouped according to their use of appropriate forecasting models. These categories are network representation, operations research, machine learning, probabilistic models, and statistical analysis. From the timeline of all articles that spot trends and relationships, the spread of delay and root delay dominate over the analysis of cancellation. Also, the researcher had previously focused on statistical

**Table 1** Studies of causal factors of flight delays using traditional methods

Research objectives	Methodology used	The result	Limitation	Resource
Identify causes of aviation delay Determine and examine approaches to reduce air traffic delays	"A comprehensive approach"	Weather factors highly affect flight delays	Many variables must be considered, like the quality and usage of multi-hour forecasts The continued analysis of the operations during various types of delay events is needed	[11]
Determine the main factors influencing flight delays	Analytical hierarchical process (AHP) to categorize the factors	The factors that affect flight delays are technical faults and delayed procedures	Did not mention	[12]

**Table 2** Studies for the flight delays' factors of flight delays by using new technologies

The objectives	Machine-learning algorithms	Result	Limitation	Researcher
Identify the periodic patterns of domestic flight arrival delays and identify the factors correlated with them	Two-stage approach	The hour, day, season, flight distance, and precipitation affect flight arrival delay most	Using the method with other airports and different climate attributes has not been tested	[13]
Improving a model to assess the distribution of flight departure delays	“Smoothing spline model”	A model is used to determine the correlation between the manner of delay distribution and seasonal trends and other effects	The model focused on the performance of United airlines in Denver international airport and needs such a model for all major US airlines and airports	[14]
Testing the performance of deep learning models in flight delay prediction	Deep learning models	For modeling sequential data, Recurrent Neural Network (RNN) has high performance	The data used has a small amount of flight data and needs big data to use deep learning models	[15]
Improving a spatial analysis approach for identifying the delay and its factors	Spatial error and lag models “SEM and SLM”	Spatial error and lag models have the best fit	Did not mention	[16]
Suggest a novel method to predict the flight departure delay	Explore supervised learning methods	The best result (high accuracy and minimum mean absolute error) was the LightGBM model	The study expected the flight departure delay from the airport and disregarded some factors, such as weather, destination, network states and so on	[17]

analysis and operational research. However, utilization of data management and machine learning is increasing.

An arrival flight delay prediction was implemented using “weather conditions” [19]. The flight information and weather status at the airports at origin and destination, depending on the flight schedule, should be considered. Parallel algorithms that executed by using a cloud platform are used to analyze and mine flight and weather

observation datasets. These algorithms result in a high accuracy for predicting flight delay.

The authors have provided a novel model for flight delay prediction with deep learning technology [19]. They tested this model on a US flight dataset, which contained noisy flight delay data. To handle the noisy data, a stack-denoising autoencoder has been used in designing a technique and added to the proposed model. Additionally, they used the “Levenberg-Marquart algorithm” to determine the appropriate weights and biases for the model. To investigate the impact of the “stack-denoising autoencoder” and “LM algorithm” on the model structure, two other models have been developed. The first model depended on the “autoencoder” and the “LM algorithm,” and the second model depended only on the “denoising autoencoder.” The authors used the confusion matrix to assess the three models in two different cases. They compared the performance of the proposed prediction model to previous methods. The results indicate that the “SDA-LM” model is better than those of the “SAE-LM” and “SDA” models. Furthermore, the accuracy of the proposed model, known as RNN, is higher than that of the previous model.

The authors of [20] introduced a new approach called “ST-Random Forest” to predict flight delays using both temporal and spatial perspectives. The data used was China flight data between June and August 2016 to test their proposed method. The authors first utilized complex network theory to extract spatial features of the aviation network at different levels (e.g., edge, network, and node). Additionally, they incorporated the temporal correlation of weather conditions and airport congestion on flight delays and developed a prediction framework using “LSTM units” to capture the temporal properties. Finally, they used Random Forest as a classifier, taking into account various factors that influence flight delays, such as spatial, temporal, and extrinsic factors. The proposed model achieved an accuracy of 92.39%, with approximately 86% of on-time samples correctly identified and 95% of delayed samples accurately classified.

The study [21] used ensemble machine-learning models to predict flight delay. Table 3 presents the comparison. The results are presented that the ensemble of “Logistic Regression (LR) + Neural Networks (NN) + Random Forest (RF)” has the highest accuracy percentage compared to other models.

The authors of [22] also used machine-learning models for flight delay prediction and compared them. They used flight data from “September 2017 to April 2023” in Saudi Arabia. The finding presented that “CatBoost” has the highest accuracy percentage compared to other selected models. The results are presented in Table 4.

**Table 3** Comparison between the selected machine-learning models and ensemble machine-learning models [21]

Model	Neural Networks (NN)	Decision tree (DT)	Random Forest (RF)	Logistic Regression (LR)	Nn + RF	LR + NN + RF
Accuracy	61.4	65.1	66.5	63.8	66.3	66.9

**Table 4** Comparison between the selected machine-learning models [22]

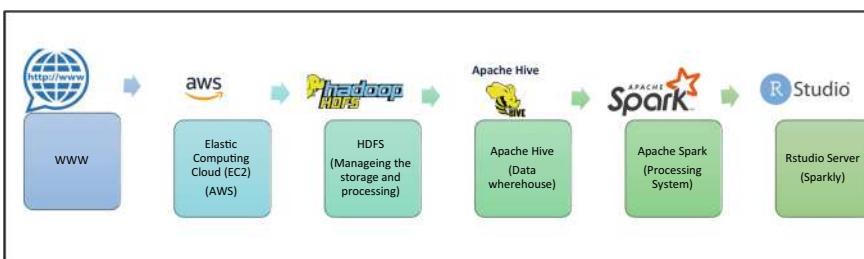
Model	CatBoost	XGBoost	LightGBM	MLP	Random Forests
Accuracy	76	73.1	73.2	72.5	71.3

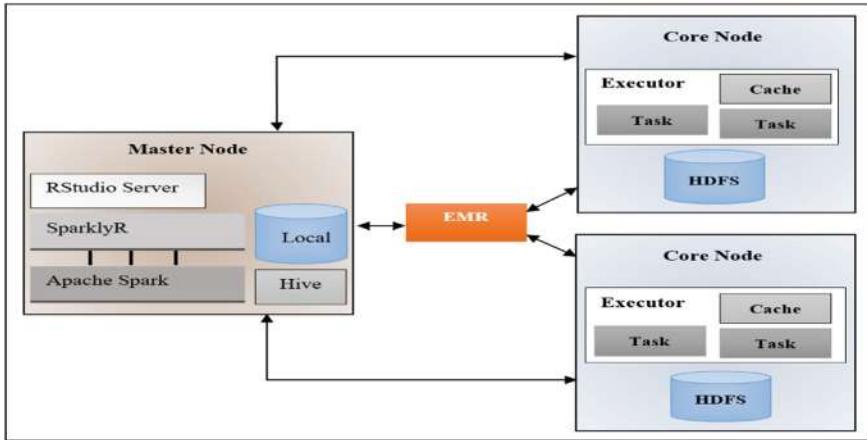
### 3 Method

This study used “big data analytics” and “business intelligence” to analyze, visualize, and predict flight delays. Big data processing needs a specific environment that can provide highly effective performance. An Apache Spark cluster and SparklyR have been selected for analysis, visualization, and prediction. Furthermore, six machine-learning models have been used for flight delay prediction: “Random Forest classifier, Logistic Regression classifier, Gradient Boosting classifier, Neural Networks classifiers, Support Vector Machine (classifier), and K-Nearest Neighbors classifier.” All models learned and tested the same dataset and features. The dataset used is available on the web. Hive tables on HDFS were used to store the data across multiple worker nodes: One controller node, which is the driver node, and two core nodes, which are the worker nodes. The RStudio Server was installed on the controller node because it is considered a coordinator of the analysis in Spark. The basic workflow for analysis, visualization, and prediction in this study is shown in Fig. 1.

#### 3.1 Setting Up the Cluster

The cluster used was provided by Amazon Web Services (AWS). To set up the cluster easily, Elastic MapReduce (EMR) was used as a cluster manager with one controller node and two core nodes so that the nodes used the virtual servers from the Elastic Compute Cloud (EC2). The architecture of the controller node and core nodes with the cluster manager, as utilized in this study, is presented in Fig. 2. The controller node is connected to the EMR through the Secure Shell Protocol (SSH). SSH required an AWS key pair (.pem key) to connect to the EC2 controller node. In addition, a

**Fig. 1** Basic workflow for analysis, visualization, and prediction in this study



**Fig. 2** Architecture of master node and core nodes with EMR

security group was necessary to give access to the RStudio Server. All software (e.g., Apache Spark, Hive, and RStudio Server) was installed on the controller node. The user directory for the RStudio user that performed the data analysis was created on a Hadoop Distributed File System (HDFS).

### 3.2 Data Source

The dataset used is flight data from the American Statistical Association. It represents nearly 29 million records of arrival and departure flight details over five years, between January 2019 and August 2023, for all commercial flights within the US. The size of the flight data is approximately 600 MB compressed and 10 GB in uncompressed CSV format, which is considered a large dataset. Table 5 shows the attributes and their values in the flight data.

## 4 Implementations

In this stage, we discussed the feature engineering and described the necessary data pre-processing step before the models commence the learning process. We also discuss the selected models and the motivation for their selection.

**Table 5** Attributes and values of used flight data

Attribute	Values	Attribute	Values	Attribute	Values
FL_DATE	2019–2023	WHEELS_OFF	Local, hhmm	DISTANCE	Miles
AIRLINE_CODE	String	WHEELS_ON	Local, hhmm	TAXI_IN	Minutes
DOT_CODE	Code	ELAPSED_TIME	Minutes	TAXI_OUT	Minutes
FL_NUMBER	Int	CRS_ELAPSED_TIME	Minutes	CANCELLED	0 or 1
DEP_TIME	Local, hhmm	AIR_TIME	Minutes	DELAY_DUE_CARRIER	Minutes
CRS_DEP_TIME	Local, hhmm	ARR_DELAY	Minutes	DELAY_DUE_WEATHER	Minutes
ARR_TIME	Local, hhmm	DEP_DELAY	Minutes	DELAY_DUE_NAS	Minutes
CRS_ARR_TIME	Local, hhmm	ORIGIN	Code	DELAY_DUE_SECURITY	Minutes
DIVERTED	0 or 1	DEST	Code	DELAY_DUE_LATE_AIRCRAFT	Minutes
ORIGIN_CITY	City name	DEST_CITY	City Name		
CANCELLATION_CODE	“A = carrier, B = weather, C = NAS, D = security, F = Late aircraft”				

## 4.1 Feature Engineering

Before training the models, the data was pre-processed. This study's five feature engineering steps are column selection, null value removal, delay filtering, column creation, and data split. A detailed explanation of each step follows.

### 4.1.1 Column Selection

Nine columns were selected for model learning, analysis, and visualization. They are “fl\_date, dep\_delay, arr\_delay, distance, delay\_due\_carrier, delay\_due\_weather, delay\_due\_nas, delay\_due\_security, and delay\_due\_late\_aircraft.” These columns were chosen because they were believed to be the most important factors in determining whether a flight would be delayed.

#### 4.1.2 Null Value Removal and Delay Filtering

Then, null values were removed from these columns to ensure the models would not be trained on incomplete data. The Arr\_Delay and Dep\_Delay columns were filtered to include results within specific ranges to remove outliers and focus on the more typical delays. The Arr\_Delay column was filtered to include results between  $-60$  and  $360$ . Similarly, the Dep\_Delay column was filtered to include results between  $-60$  and  $240$ .

#### 4.1.3 Column Creation

Five new columns were created: “Year, month, day of month, gain, and delayed.” The year, month, and day of month columns are derived from fl\_date columns. The “gain” column results from subtracting the Dep\_Delay value and the Arr\_Delay value; that is,  $\text{gain} = \text{Dep\_Delay} - \text{Arr\_Delay}$ . The “delayed” column is a binary column that indicates whether the flight was delayed. Its values are  $0 = \text{“not delayed”}$  and  $1 = \text{“delayed.”}$  This column was created using an if-else statement: If  $\text{gain} >= 15$ , then the value will equal  $1$ ; otherwise, the value will be equal to  $0$ .

#### 4.1.4 Data Split

The pre-processed data was split into a “training set” and a “test set,” with “70%” of the data in the “training set” to train the models and “30%” in the “test set” to evaluate the models’ performance.

### 4.2 *Models*

The problem of flight delay prediction in this study is considered a classification problem because there are only two possible outcomes: Delayed or not delayed.

#### 4.2.1 Selected Algorithms

The selected algorithms for this problem were “Random Forest classifier, Logistic Regression classifier, Neural Networks classifiers, Support Vector Machine (classifier), Gradient Boosting classifier, and K-Nearest Neighbors classifier.” The six selected algorithms are all classification algorithms that effectively predict flight delays and are used for big data and normal-sized datasets.

#### 4.2.2 Motivation for the Models' Selection

The authors reviewed related studies to see which algorithms had been used for flight delay prediction. They found that most studies used at least one or two of the six algorithms selected in this study. Therefore, the authors selected the algorithms depending on the results of the related studies. Another reason for choosing these algorithms was that this study had to select a set that differed from those chosen in the related studies.

Random Forest is “an ensemble technique proposed by Breiman for classification problems. It boosts the system’s accuracy by combining several models to solve the problem” [23]. Utilizing several decision models often leads to less precise predictions than using a single model. This model is the most suitable machine-learning technique for classification problems in various study disciplines since it can utilize training data from randomly selected subsets and create trees in a similarly random manner [22, 23]. Logistic Regression classifier is “a widely used statistical model that allows for multivariate analysis and modeling of a binary dependent variable. It is a similar model for a continuous dependent variable” [24]. The motivations for using Neural Network classifiers are their ability to handle noisy data effectively and to classify patterns even in situations where they have not explicitly been trained. The motivations for using Support Vector Machines are their effectiveness in dealing with high-dimensional spaces and when the number of dimensions is greater than the number of patterns, as well as their efficient use of memory. A “gradient-boosted tree” is a method used to further develop regression and classification models and other models’ learning processes [25]. These models are often non-linear and known as regression or decision trees. Adding new learners gradually and sequentially is how a group of weak prediction models, such as “regression decision trees,” can be modeled. The structure of a new learner is composed of nodes and leaves, leading to the prediction results depending on the decision nodes. Gradually, the ensembles are built up, and every newly formed ensemble corrects the faults in the previous ensemble. K-Nearest Neighbors classifier is used because of its simplicity. Its process depends on classifying the new data points according to the similarity measure of the earlier stored data points.

## 5 Result and Discussion

In the following discussion, the flight data analyzed depends on multiple points, such as the total number of delayed flights by reason, the total number of delayed and on-time flights per year, the total number of delayed or on-time flights per month, and so on. Then, the results of the flight data analysis have been visualized. After the models’ learning and testing, all models’ performance was measured by F1-score, accuracy, precision, and recall.

**Table 6** Total number of delayed flights by reason

Reasons	Number of flights	Percentage
On-time	24,788,794	0.7927
National Aviation System Delay	1,530,541	0.0489
Security delay	16,407	0.0005
Aircraft Arriving Late	1,921,449	0.0614
Air Carrier Delay	1,893,614	0.0606
Weather delay	210,015	0.0067

## 5.1 Flight Data Analysis and Visualization

In the flight data analysis and visualization, the causal factors of delayed flights can be identified. Some points can help the analyst to identify the causal factors of flights delayed, including the total number of delayed flights, the total number of delayed flights by reason, the total number of delayed or on-time flights per year, the total number of delayed or on-time flights per month, the total number of delayed or on-time flights per day of the week, and the total number of delayed or on-time flights per day of the month. The analysis and visualization in this step are applied to all arrival and departure flights.

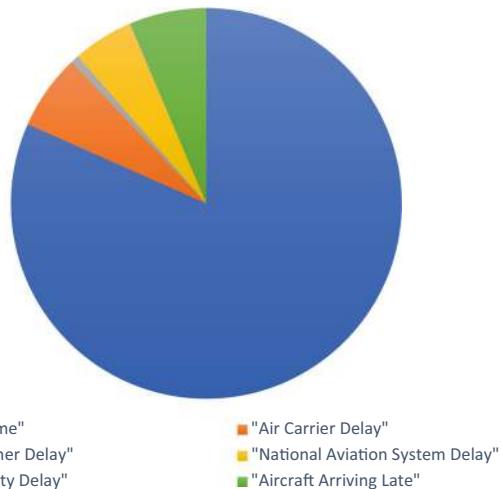
### 5.1.1 Total Number of Delayed Flights by Reason

In the flight data, the reasons for delayed flights were classified into five reasons: “Weather delay, NAS delay, security delay, late aircraft delay, and carrier delay.” Weather conditions cause a weather delay. A “NAS delay” is within the control of the NAS, including airport operations, heavy traffic, and air traffic control. A security threat causes security delays. An arrival delay at an airport causes a late aircraft delay due to the late arrival of the same aircraft at a previous airport. A carrier delay is within the control of the air carrier. Table 6 and Fig. 3 present the total number of on-time and delayed flights by reason. The results show that Aircraft Arriving Late and Air Carrier Delay reasons have the highest percentage compared to other reasons, followed by National Aviation System Delay and weather delay. The security delay has the lowest percentage.

### 5.1.2 Total Number of Delayed or On-Time Flights per Year

In this step, the total number of delayed and on-time flights for all commercial flights within the US are analyzed according to years. Table 7 and Fig. 4 show the total number of delayed arrivals and departures and on-time flights per year. The

**Fig. 3** On-time flights and delayed flights by reason



**Table 7** Total number of delayed and on-time flights per year

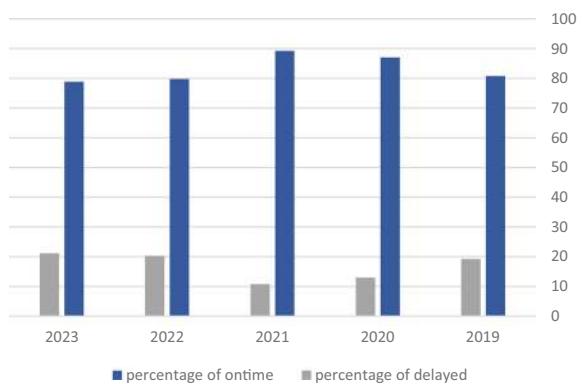
Year	On-time arrivals	On-time departures	Arrival delays	Departure delays	Total on-time flights	Total delayed flights	Percentage of on-time	Percentage of delayed
2019	1,488,204	1,512,007	366,802	347,365	3,000,211	714,167	80.7729	19.2271
2020	1,604,335	1,616,297	244,465	235,849	3,220,632	480,314	87.02186	12.97814
2021	1,031,503	1,048,909	132,794	117,511	2,080,412	250,305	89.2606	10.7394
2022	1,278,339	1,280,843	324,158	325,192	2,559,182	649,350	79.76177	20.23823
2023	1,327,452	1,348,027	366,312	349,748	2,675,479	716,060	78.88687	21.11313

results show that 2023 has the highest number of delayed flights, followed by 2022, then 2019, and 2020. Finally, 2021 has the lowest number of delayed flights.

### 5.1.3 The Mean of Delayed Flights per Month

In this step, the mean delayed flights are analyzed by month. Table 8 and Fig. 5 present the mean of monthly delayed flights. The results show that June has the highest number of delayed flights, followed by July. September and November have the lowest number of delayed flights.

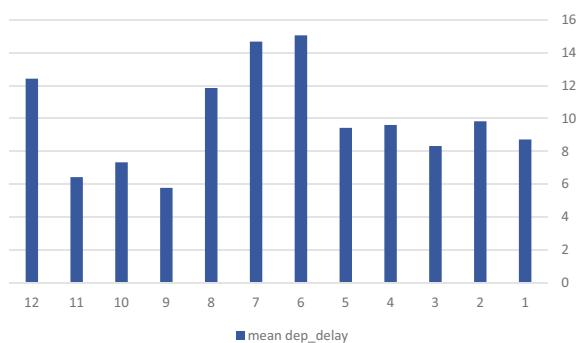
**Fig. 4** Total number of delayed and on-time flights per year



**Table 8** Mean of delayed flights per month

Month	Mean dep_delay
1	8.718996
2	9.823191
3	8.320393
4	9.597162
5	9.425539
6	15.047865
7	14.666483
8	11.84619
9	5.776012
10	7.330811
11	6.430303
12	12.413276

**Fig. 5** Mean of delayed departure and arrival flights per month



## 5.2 Calculating the Correlation Matrix

The correlation matrix has been calculated between eleven columns: Dep\_delay, year, Month, day of month, arr\_delay, distance, carrier\_delay, weather\_delay, NAS\_delay, security\_delay, and late\_aircraft\_delay. Table 9 presents the correlation matrix between the selected features.

The result shows that the Dep\_delay column has the most significant relationship with Arr\_delay column compared to the other features, followed by carrier\_delay, late\_aircraft\_delay, weather\_delay, NAS\_delay, year, security\_delay, distance, and day of month. The month column has the least significant relationship with the Dep\_delay and Arr\_delay columns.

## 5.3 Models' Evaluation

F1-score, accuracy, precision, and recall were used to evaluate the models' performance. The results are shown in Table 10. The results of each metric have been explained in the following:

- **F1-score:** The results show that the Gradient Boosting classifier and K-Nearest Neighbors' classifier are equal to 0.92, the Logistic Regression classifier is equal to 0.82 (which is the lowest result), the Neural Network classifier is equal to 0.94, the Random Forest classifier is equal to 0.93, and the Support Vector Machine is equal to 0.99 (which is the best result).
- **Accuracy:** The results show that the Logistic Regression classifier and the Gradient Boosting classifier are equal to 0.93, the K-Nearest Neighbors' classifier is equal to 0.88, the Neural Network classifier is equal to 0.89, the Random Forest classifier is equal to 0.87 (which is the lowest result), and the Support Vector Machine is equal to 0.98 (which is the best result).
- **Recall:** The results show that the Random Forest classifier and K-Nearest Neighbors Classifier are equal to 0.9, the Logistic Regression classifier is equal to 0.7 (the lowest result), and the Neural Network classifier, Support Vector Machine, and Gradient Boosting classifier are equal to 0.99 (the best results compared to other models).
- **Precision:** The results show that the Random Forest classifier is equal to 0.96, the K-Nearest Neighbors classifier is equal to 0.95, the Logistic Regression classifier is equal to 0.99 (the best result compared to other models), the Neural Network classifier is equal to 0.9, the Support Vector Machine is 0.98, and Gradient Boosting classifier is equal to 0.85 (the lowest results compared to other models).

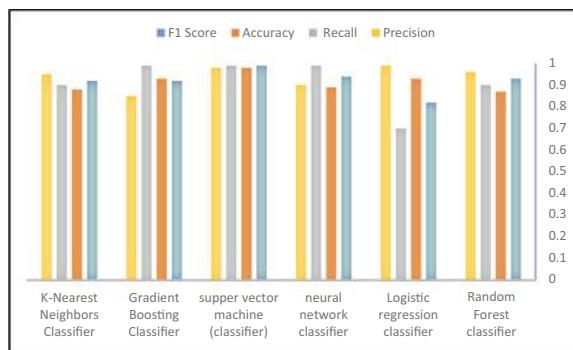
Figure 6 presents the performance results of six machine-learning models. The results of the models' performance present that the results of Random Forest classifier and K-Nearest Neighbors' classifier are approximately equal. Finally, the results

**Table 9** Heatmap of correlation matrix between the selected features

	Dep_delay	Year	Month	Day of month	Arr_delay	Distance	Carrier_delay	Weather_delay	NAS_delay	Security_delay	Late_aircraft_delay
Dep_delay	1	0.04	0.009	0.01	0.97	0.02	0.68	0.31	0.28	0.035	0.6
Year	0.04	1	0.12	0	0.03	0.03	0.143	0.011	0.3	0.02	0.38
Month	0.009	0.12	1	0.01	0	0	0.27	0.0411	0.0212	0.01	0.16
Day of month	0.01	0	0.01	1	0.01	0	0.02	0.22	0.05	0.001	0.004
Arr_delay	0.97	0.03	0	0.01	1	0	0.67	0.33	0.37	0.035	0.59
Distance	0.02	0.03	0	0	0	1	0.012	-0.008	0.019	0.004	-0.008
Carrier_delay	0.68	0.143	0.27	0.02	0.67	0.012	1	-0.004	0.013	-0.001	0.051
Weather_delay	0.31	0.011	0.0411	0.22	0.33	-0.008	-0.004	1	0.023	0	0.02
NAS_delay	0.28	0.3	0.0212	0.05	0.37	0.019	0.013	0.023	1	0.002	0.039
Security_delay	0.035	0.02	0.01	0.001	0.035	0.004	-0.001	0	0.002	1	0.003
Late_aircraft_delay	0.6	0.38	0.16	0.004	0.59	-0.008	0.051	0.02	0.039	0.003	1

**Table 10** Models' performance results

Model	F1-score	Accuracy	Recall	Precision
Random Forest classifier	0.93	0.87	0.9	0.96
Logistic Regression classifier	0.82	0.93	0.7	0.99
Neural Network classifier	0.94	0.89	0.99	0.9
Support Vector Machine (classifier)	0.99	0.98	0.99	0.98
Gradient Boosting classifier	0.92	0.93	0.99	0.85
K-Nearest Neighbors classifier	0.92	0.88	0.9	0.95

**Fig. 6** Comparison between the six models' performance measurement results

showed that the Support Vector Machine is the best model and the Logistic Regression classifier is the worst model, but it is still comparable.

There may be some limitations in this study. The desired dataset was flight data in Saudi Arabia, but acquiring the data took a long time. To avoid this limitation, the data must be acquired before conducting the study in enough time.

## 6 Conclusion

This study selected nine columns for analysis by machine-learning models. In addition, five new columns were created: "Year, month, day of month, gain, and delayed." The analysis and visualization step is applied to all commercial flights arriving and departing from January 2019 to August 2023 within the US. Aircraft Arriving Late and Air Carrier Delay reasons have the highest percentage compared to other reasons. The 2023 year had more delays than the other years. June has the highest number of delayed flights compared to other months, followed by July. The performance of the selected machine-learning models, including the Random Forest classifier, Logistic Regression classifier, Neural Networks classifiers, Support Vector Machine (classifier), Gradient Boosting classifier, and K-Nearest Neighbors classifier, were

measured. The results were presented regarding F1-score, accuracy, recall, and precision. It was found that all results for the Random Forest classifier and K-Nearest Neighbors' classifier are approximately equal. The Support Vector Machine classifier is the best model, and the Logistic Regression classifier is the worst model, but it is still comparable.

This conclusion works for flight data with the attributes and machine-learning models used. For future work, different artificial intelligence techniques, such as deep learning or the same machine-learning technique with different classification models or different attributes and conditions, can be used. The study will also be applied to flight data in Saudi Arabia.

## References

1. Sun Z, Zou H, Strang K (2015) Big data analytics as a service for business intelligence. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)* 9373:200–211. [https://doi.org/10.1007/978-3-319-25013-7\\_16/FIGURES/3](https://doi.org/10.1007/978-3-319-25013-7_16/FIGURES/3)
2. Sun Z, Sun L, Strang K (2016) Big data analytics services for enhancing business intelligence. *J Comput Inf Syst*. <https://doi.org/10.1080/08874417.2016.1220239>
3. Lim EP, Chen H, Chen G (2013) Business intelligence and analytics. *ACM Trans Manag Inf Syst* 3(4). <https://doi.org/10.1145/2407740.2407741>
4. Fan S, Lau RYK, Zhao JL (2015) Demystifying big data analytics for business intelligence through the lens of marketing mix. *Big Data Res.* 2(1):28–32. <https://doi.org/10.1016/J.BDR.2015.02.006>
5. Gandomi A, Haider M (2015) Beyond the hype: big data concepts, methods, and analytics. *Int J Inf Manage* 35(2):137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
6. Sun Z, Sun L, Strang K (2018) Big data analytics services for enhancing business intelligence. *J Comput Inf Syst* 58(2):162–169. <https://doi.org/10.1080/08874417.2016.1220239>
7. Peng ELIM, Chen H, Chen G, Peng E (2013) Part of the databases and information systems commons, and the numerical analysis and scientific computing commons citation citation. *ACM Trans Manag Inf Syst* 3(4):1–10. <https://doi.org/10.1145/2407740.2407741>
8. Gronwald KD (2017) Integrated business information systems: a holistic view of the linked business process chain ERP-SCM-CRM-BI-big data. Springer, Berlin Heidelberg
9. Yazdi MF, Kamel SR, Chabok SJM, Kheirabadi M (2020) Flight delay prediction based on deep learning and Levenberg-Marquart algorithm. *J Big Data* 7(1):1–28. <https://doi.org/10.1186/S40537-020-00380-Z/TABLES/7>
10. Zoutendijk M, Mitici M (2021) Probabilistic flight delay predictions using machine learning and applications to the flight-to-gate assignment problem. *Aerosp* 8(6):152 (2021). <https://doi.org/10.3390/AEROSPACE8060152>
11. Allan SS, Beesley JA, Evans JE, Gaddy SG (2001) Analysis of delay causality at Newark international airport. Accessed: 28 March 2021. [Online]. Available: <http://www.ll.mit.edu/AviationWeather>
12. Kazemi Asfe M, Jangi Zehi M, Shahiki Tash MN, Yaghoubi NM (2014) Ranking different factors influencing flight delay. *Manag Sci Lett* 4(7):1397–1400. <https://doi.org/10.5267/j.msl.2014.6.030>
13. Abdel-Aty M, Lee C, Bai Y, Li X, Michalak M (2007) Detecting periodic patterns of arrival delay. *J Air Transp Manag* 13(6):355–361. <https://doi.org/10.1016/j.jairtraman.2007.06.002>
14. Tu Y, Ball MO, Jank WS (2008) Estimating flight departure delay distributions—a statistical approach with long-term trend and short-term pattern. *J Am Stat Assoc* 103(481):112–125. <https://doi.org/10.1198/016214507000000257>

15. Kim YJ, Choi S, Briceno S, Mavris D (2016) A deep learning approach to flight delay prediction. In: AIAA/IEEE digital avionics systems conference—proceedings, vol 2016-Decem. <https://doi.org/10.1109/DASC.2016.7778092>
16. Cheng S, Zhang Y, Hao S, Liu R, Luo X, Luo Q (2019) Study of flight departure delay and causal factor using spatial analysis. J Adv Transp 2019. <https://doi.org/10.1155/2019/3525912>
17. Ye B, Liu B, Tian Y, Wan L (2020) A methodology for predicting aggregate flight departure delays in airports based on supervised learning. Sustain 12(7). <https://doi.org/10.3390/su12072749>
18. Sternberg A, Soares J, Carvalho D, Ogasawara E (2017) A review on flight delay prediction. Transp Rev. <https://doi.org/10.1080/01441647.2020.1861123>
19. Belcastro L, Marozzo F, Talia D, Trunfio P (2016) Using scalable data mining for predicting flight delays. ACM Trans Intell Syst Technol 8(1). <https://doi.org/10.1145/2888402>
20. Li Q, Jing R (2022) Flight delay prediction from spatial and temporal perspective. Expert Syst Appl 205:117662. <https://doi.org/10.1016/J.ESWA.2022.117662>
21. Chaitanya GS, Shaik DS, Visalakshi P, Rakshitha G (2024) Comparative study of ensemble machine learning techniques for airline delay prediction. Int J Eng Comput Sci 6(1):14–21. <https://doi.org/10.33545/26633582.2024.V6.I1A.105>
22. Alfarhood M, Alotaibi R, Abdulrahim B, Einieh A, Almousa M, Alkhanifer A (2024) Predicting flight delays with machine learning: a case study from Saudi Arabian airlines. Int J Aerosp Eng 2024(1):3385463. <https://doi.org/10.1155/2024/3385463>
23. Shaheed K, Szczuko P, Abbas Q, Hussain A, Albathan M (2023) Computer-aided diagnosis of COVID-19 from chest X-ray images using hybrid-features and random forest classifier. Healthcare 11(6):837. <https://doi.org/10.3390/HEALTHCARE11060837>
24. Shipe ME, Deppen SA, Farjah F, Grogan EL (2019) Developing prediction models for clinical use using logistic regression: an overview. J Thorac Dis 11(Suppl 4):S574. <https://doi.org/10.21037/JTD.2019.01.25>
25. Hengl T, Nussbaum M, Wright MN, Heuvelink GBM, Gräler B (2018) Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 2018(8):e5518. <https://doi.org/10.7717/PEERJ.5518/SUPP-1>

# A Location-Specific Mobile Framework for Intelligent Road Traffic Traceability Systems



**Khushi Saxena, Tiansheng Yang, Ruikai Sun, Changgui Lin, Lu Wang, and Rajkumar Singh Rathore**

**Abstract** The amalgamation of the Internet of Things (IoT) technology into Intelligent Traffic Monitoring Systems (ITMS) has changed the way we can handle, manage, and analyze traffic data. Slowly and steadily, the globe is getting more and more connected, so the effect on different traffic control and monitoring systems by IoT majorly influences the vitality and impact. This research paper focuses on the complex association between IoT and Intelligent traffic monitoring systems, contributing a detailed study of how technologies are converting and altering our monitoring and analyzing the traffic movement specifically in IoT. After thorough research and examination of IoT, we uncover various strengths, weaknesses, opportunities, threats, and future scopes to help promote IoT. Through this study, we will be able to enlighten and even contribute to the impactful ways of incorporating IoT in traffic management systems by using sensors and transmitters to better conveyance systems.

---

K. Saxena

Kalinga Institute of Industrial Technology, Bhubaneswar, India

e-mail: [22054046@kiit.ac.in](mailto:22054046@kiit.ac.in)

T. Yang (✉)

University of South Wales, Pontypridd, UK

e-mail: [tiansheng.yang1@southwales.ac.uk](mailto:tiansheng.yang1@southwales.ac.uk)

R. Sun

Cardiff University, Cardiff, UK

e-mail: [sunr10@cardiff.ac.uk](mailto:sunr10@cardiff.ac.uk)

C. Lin

Gansu Agricultural University, Anning District, Lanzhou, Gansu Province, China

e-mail: [changgui-lin@hsgylgf.com](mailto:changgui-lin@hsgylgf.com)

L. Wang

Xi'an Jiaotong-Liverpool University, Wuzhong District, Suzhou, China

R. S. Rathore

Cardiff School of Technologies, Cardiff Metropolitan University, Cardiff, UK

e-mail: [rsrathore@cardiffmet.ac.uk](mailto:rsrathore@cardiffmet.ac.uk)

**Keywords** Internet of Things (IoT) · Intelligent traffic monitoring systems (ITMS) · Sensors · Transmitters

## 1 Introduction

The Internet of Things (IoT) has helped with the supervision of traffic framework in several methods including ordered precise concurrent information analyzed statistics and sped-up decision-making by the introduction of sensors and transmitters among basic objects that can ease out the labor and extensive efforts to fetch the data and scrutinize it by adding deep learning machine learning and more to minimize the anomalies and inconsistency. These nodes and objects having IoT features in them not only help with management but also support concurrent supervision efficiency and penny-saving provision for users. These models are based on IoT which assists us particularly with data analysis and acquisition by extracting important statistics and then streamlining the details being gathered. Data warehouse and optimization have helped with effective data archival and optimizing data retention. Lastly, real-time traffic updates provide accurate and timely information to users along with dynamic traffic notifications and even instant alerts. This will help us to minimize traffic flow and enhance basic facilities for conveyance of the general public. Moving on to the utility of IoT and ITMS, commuters can have smooth transportation, optimized traffic flow, alleviated traffic jams, maintained traffic flow, streamlined traffic movement, mitigated delays, fewer disruptions, enhanced mobility, and improved vehicular navigation and movement. The association of Internet of Things (IoT) technology with Intelligent Traffic Monitoring Systems indicates evidential progress in the field of conveyance management. The consequences of IoT on these systems are indisputable, as they have possibilities to overturn the way traffic data is accumulated, refined, and used. By capitalizing on and taking advantage of IoT technology, intelligent traffic monitoring systems can not only enhance traffic flow but also enhance passenger comfort by improving journeys and making trips pleasant, reducing travel interruptions by reducing stoppages and reducing transit time, and optimizing travel flow by ensuring smooth traffic flow and more. The all-encompassing analysis conferred in this study marks the metamorphic ability of IoT-enabled systems to address critical issues in traffic management. Nevertheless, regardless of the favorable outputs highlighted in this research paper, it is important to acknowledge and admit possible drawbacks or gaps in the research. The future scope could lean toward enhancing long-term sustainability and dependability as well as betterment in the investigation of public privacy and security which is related to the data analysis, acquisition, and storage. An illustration of the traffic control and tracking framework is depicted in Fig. 1.

The Internet of Things has introduced us to a new and advanced age full of internally connected intricately-detailed novelty and impactful changes in our day-to-day lives among several technologies IoT and ITMS have been some of the most

**Fig. 1** Illustration of traffic control and tracking framework



prominent and remarkable technologies playing exceptional roles in the incorporation of these techniques that can assist us bring huge impacts to the world like having smooth flow and control of congestion administration framework by involving sensors transmitters digital nodes and more promoting durability have better productivity, cost-effectiveness, and even welfare of individuals. This research paper gives a detailed explanation of IoT along with Intelligent Traffic Management Systems involving conveyance surveillance framework control platforms incident detection and congestion management platforms along with several sharp mobility resolutions. It also brings light to the future scope and opportunities in conveyance for both rural and city areas by investigating and analyzing real-life scenarios having cutting-edge and revolutionary innovations and implementing multidimensional policy ramifications and changes from different perspectives to life. Finally, this paper also helps the urban and rural city and policy planners to learn from the amalgamation of IoT and Intelligent Traffic Management Systems while providing views and ideas for better traffic transit management.

## 2 Background Study

In [1], the model that is proposed is about reducing congestion to have much better road organization and safety so that emergency conveyance like ambulances can travel within time and wouldn't have to stand stuck in traffic.

In [2], here not just theory and details are provided but also visual representation consisting of graphs and pie chart, to collect varied yet accurate information about the congestion on single handedly using ITMS. In [3], here we use RSU which is the Road Side Unit and ITMS together to implement better usage of sensors, transmitters, database and enhance ease in browsing traffic related information. In [4], here while in traffic we can detect the amount of time it will take to get out of it and this is done with the help of PEPA, TOPSIS, and IoT. Where PEPA plays an important role in enhancing performance and provides instantaneous traffic status. In [5], here RSU is

implemented to capture images of the conveyance and getting information about it instantly while utilizing IoT and ITMS. In [6], here features like GPS, and RFID are made with the help of IoT, Cloud Computing and Machine Learning to make vehicle information and location tracking concurrent and real time with great speed. In [7], here techniques like IoT, ITMS and Machine learning are used for the ease, accuracy and efficiency features that are to be induced in vehicle congestion, so as to enhance emergency detection. In [8], here many techniques like IoT and Machine Learning have been used for distributes multiagent Q—learning. Lastly, this will make both people with and without vehicles have a much optimized and safe travel.

### 3 Proposed Model

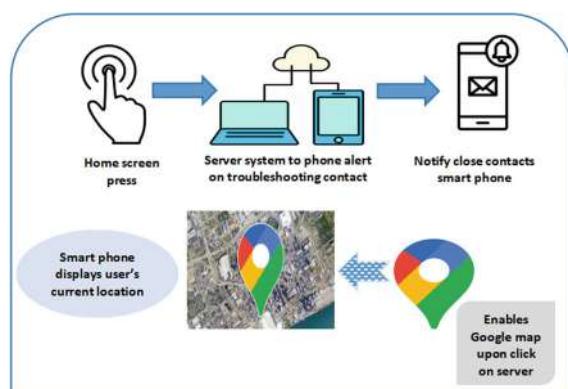
The Intelligent Traffic Monitoring System (ITMS) is an advanced application of (IoT) Internet of Things that enhances congestion control, efficiency, and safety in urban areas. The key elements functions and advantages of an ITMS are depicted in this model along with potential uses and future developments [9]. Effective traffic management has become a serious issue, in the busy urban areas impacting not only the lives of people, but also the economic productivity and environmental sustainability of cities not only in India, but worldwide. There is an increasing need for innovative solutions such as traffic incidents and congested roads continue to pose safety risks and dangers. The ITMS is a highly developed interface used to resolve the real-life problems associated with existing congestion management with unparalleled efficiency and effectiveness. ITMS takes advantage of the potential of the Internet of Things by deploying a series of sensors, cameras, data analytics algorithms throughout cities [10]. They have sensors that capture and transmit real-time data about vehicle movements traffic flow etc. They are placed at such intersections and routes, that it is easy to achieve data efficiently owing to the volume of data the information gathered and kept in the system by ITMS enabling it to dynamically monitor, supervise, and optimize congestion operations. The ITMS is not just limited to gathering information but also helps with concurrent decision-making by congestion authorities and city planners is now made possible by these technologies and their ability to extract notable views from enormous amounts of congestion data where, transformative technologies like machine learning and even artificial intelligence help out with algorithms. Moreover, there is more than just a traffic control device, it is a catalyst for larger common benefits such as decreasing traffic cutting down on travel times and enhancing safety on roads the system increases the standard of living for people living in urban areas [11–15].

A smart and intelligent application can be developed to keep track of real-time traffic since ITMS helps manage traffic. One of the main reasons that cause hectic traffic jams are accidents taking place mid-road while people make a crowd around it, and it takes time for the police and others to arrive to manage the accident and crowd. So introducing this application will help us effectively reduce accidents and waiting time of others clearing the roads quicker. The application will be based on

two major principles—Safety and Connectivity. The location will be known instantly and a notification will be sent right away. It will guarantee safety and security at all times from identifying traffic congestion and accidents to optimizing traffic signal timings and routing. We can provide stakeholders, with the tools they need to manage traffic flows and avoid bottlenecks. Hence, an Intelligent Traffic Monitoring System application, that is through and guides in the design and implementation of IoT technology. This kind of system will be able to deliver real-time data, analytics, plus an advanced communication infrastructure, which could radically transform urban transportation systems toward improving safety levels as well as elevating quality standards for people living in cities [16–18]. Figure 2 depicts the applications UI which provides a more visual representation and allows users to simply understand and know their alternatives as well as transmit distress signals when necessary. This will include categories like SOS, sending or receiving other peoples positions maps the closest alternative, or probable path profile and more the influence of IoT on Intelligent Traffic Monitoring Systems, reveals several noteworthy findings to begin Internet of Things technologies has the ability to significantly enhance the efficiency and efficacy with regard to the traffic management systems by allowing for proactive interventions and adaptive reactions to changing traffic circumstances, concurrent data, acquisition, and utilization allows transportation authorities. In terms of discovering congestion, hotspots utilize and minimize signal timings and improve incident management resulting in smoother traffic flows and shorter travel times. The report does however identify obstacles and issues that must be addressed in order to completely understand the advantages of Internet of Things, in congestion monitoring such as privacy concerns, cybersecurity risks, and the need for solid architecture and data governance frameworks. Altogether, the analysis emphasizes the significance of strategic planning collaboration and speculations in IoT-enabled outputs for addressing the complex challenges of city movement and transportation in the twenty-first century.

Here, first to send a distress signal, we need to click on the home screen for approximately 3 s then the signal will be sent to the server computer and mobile by a

**Fig. 2** Procedure of proposed model



notification and the helpline number will be called. The notification will also be sent to the people nearby the user's phone and a few selected contact members so they can help. Then after clicking on the notification, the google maps will open and with the help of API the users location will be located and help can be further sent. Live location will be available to the contacts selected, people nearby and the authorities. According to the situation, the police and higher authorities can be informed.

## 4 Results and Discussion

The experimental set up is undertaken in 5 varying phases, and the information details of vehicles at a busy traffic scene is taken into consideration. Using python, the model is simulated and its outcome is depicted in this section.

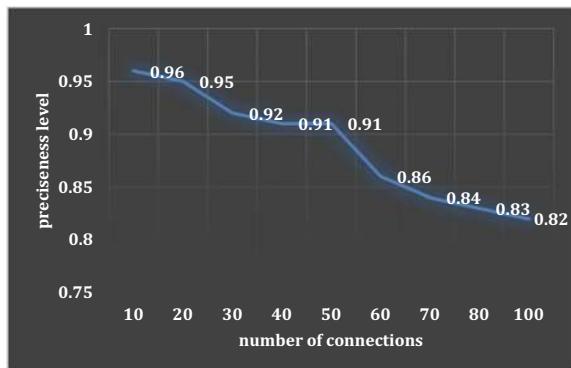
A mean response time analysis is carried out to show the model's latency period in real time. For comparison purpose, 7 successive weeks are evaluated, and it is observed that the response time fluctuates over different weeks. Response time for fourth week is 1.23 s which is found to be least while a response time of 1.56 s is the highest value obtained. The mean response time is computed to be 1.36 s only. Figure 3 highlights the overall procedure.

Figure 4 discusses the dissection between number of connections and the preciseness level of the proposed model. This is done to validate the scalability of the model when more connections are deployed. From the analysis, it is observed that the preciseness level remains stable as long as the connections are below 50. When it is 100 then the preciseness level dips to 0.82. Still the level is satisfactory, as it is not diminishing drastically.

**Fig. 3** Response time analysis with regards to the proposed model



**Fig. 4** Scalability analysis as described by the proposed model



## 5 Conclusion

The overall impact and rise with regard to modern traffic regulation system is highlighted in this study. A novel smart and intelligent traffic management model is presented here using smart phone-based location-specific information control. The model when implemented generated promising results. The mean response time is 1.36 s while the scalability testing is also done in context to number of connections. The outcome is positive and is definitely recommended for societal cause especially in urban zones with heavy traffic.

## References

- Bhate SV, Kulkarni PV, Lagad SD, Shinde MD, Patil S (2018) IoT based intelligent traffic signal system for emergency vehicles. In: 2018 second international conference on inventive communication and computational technologies (ICICCT), pp 788–793
- Chowdhury A (2016) Priority based and secured traffic management system for emergency vehicle using IoT. In: 2016 international conference on engineering and MIS (ICEMIS), pp 1–6
- Ding J, Wang R, Chen X (2016) Performance modeling and evaluation of real-time traffic status query for intelligent traffic systems. In: 2016 22nd Asia-Pacific conference on communications (APCC), pp 238–242
- Egea S, Rego Manez A, Carro B, Sanchez Esguevillas A, Lloret J (2018) Intelligent IoT traffic classification using novel search strategy for fast based-correlation feature selection in industrial environments. *IEEE Internet Things J* 5(3):1616–1624
- Latif S, Afzaal H, Zafar NA (2018) Intelligent traffic monitoring and guidance system for smart city. In: 2018 international conference on computing, mathematics and engineering technologies (iCoMET), pp 1–6
- Liu Y, Liu L, Chen W-P (2017) Intelligent traffic light control using distributed multi-agent Q learning. In: 2017 IEEE 20th international conference on intelligent transportation systems (ITSC), pp 1–8
- Pendor RB, Tasgaonkar PP (2016) An IoT framework for intelligent vehicle monitoring system. In: 2016 international conference on communication and signal processing (ICCP), pp 1694–1696

8. Pyykonen P, Laitinen J, Viitanen J, Eloranta P, Korhonen T (2013) IoT for intelligent traffic system. In: 2013 IEEE 9th international conference on intelligent computer communication and processing (ICCP), pp 175–179
9. Sharma V, Mahanayak SP, Thapa T, Mishra S, Alkhayyat A (2023) Leveraging the synergy of edge computing and IoT in supply chain management. In: 2023 10th IEEE Uttar Pradesh section international conference on electrical, electronics and computer engineering (UPCON), vol 10. IEEE, pp 1055–1062
10. Ranjan H, Mishra S, Al-Khasawneh MA, Singhal M, Sharma V, Alkhayyat A (2023) KESMR: a knowledge enrichment semantic model for recommending microblogs. In: 2023 10th IEEE Uttar Pradesh section international conference on electrical, electronics and computer engineering (UPCON), vol 10. IEEE, pp 669–674
11. Dimri SC, Indu R, Bajaj M, Rathore RS, Blazek V, Dutta AK, Alsubai S (2024) Modeling of traffic at a road crossing and optimization of waiting time of the vehicles. *Alex Eng J* 98:114–129
12. Kumar M, Kumar S, Kashyap PK, Aggarwal G, Rathore RS, Kaiwartya O, Lloret J (2022) Green communication in internet of things: a hybrid bio-inspired intelligent approach. *Sensors* 22(10):3910
13. Bhawana, Kumar S, Rathore RS, Mahmud M, Kaiwartya O, Lloret J (2022) BEST—blockchain-enabled secure and trusted public emergency services for smart cities environment. *Sensors* 22(15):5733
14. Rathore RS, Hewage C, Kaiwartya O, Lloret J (2022) In-vehicle communication cyber security: challenges and solutions. *Sensors* 22(17):6679
15. Khasawneh AM, Singh P, Aggarwal G, Rathore RS, Kaiwartya O (2022) E-mobility advisor for connected and autonomous vehicles environments. *Adhoc Sens Wirel Netw* 53
16. Rathore RS, Kaiwartya O, Qureshi KN, Javed IT, Nagmeldin W, Abdelmaboud A, Crespi N (2022) Towards enabling fault tolerance and reliable green communications in next-generation wireless systems. *Appl Sci* 12(17):8870
17. Rathore RS, Sangwan S, Kaiwartya O (2021) Towards trusted green computing for wireless sensor networks: multi metric optimization approach. *Adhoc Sens Wirel Netw* 49
18. Saleh A, Joshi P, Rathore RS, Sengar SS (2022) Trust-aware routing mechanism through an edge node for IoT-enabled sensor networks. *Sensors* 22(20):7820

# Toward Efficient Multi-attribute Prediction: Lessons from Political Bias Detection



Charan Ramtej Kodi, Satya Sai Bharath Vemula, Nikhil Kumar Pulipeta, Varsha Venkata Krishnan, Vishal Rajkumar, and Bharath Goud Musalaya

**Abstract** The use of joint models in predicting multiple attributes simultaneously has been shown to improve performance and reduce training time when correctly modeled. For example, predicting political bias and other attributes such as the topic and source of the news can be done more efficiently and accurately through joint modeling. However, the selection of which attributes to model together, the model architecture, and the choice of text representation are all important factors that can affect the performance of such models. In this work, we study and demonstrate the importance of these factors on a political bias dataset, showing how changing the second attribute being predicted, the model architecture, and the learning representation can impact the performance of bias prediction. This research has important implications for the development of more efficient and accurate models for predicting multiple attributes in news data.

---

C. R. Kodi (✉)  
University of Hyderabad, Hyderabad, Telangana, India  
e-mail: [21mcpc10@uohyd.ac.in](mailto:21mcpc10@uohyd.ac.in)

S. S. Bharath Vemula · V. V. Krishnan · V. Rajkumar  
Purdue University, West Lafayette, IN, USA  
e-mail: [vsatysa@purdue.edu](mailto:vsatysa@purdue.edu)

V. V. Krishnan  
e-mail: [venka104@purdue.edu](mailto:venka104@purdue.edu)

V. Rajkumar  
e-mail: [rajkumav@purdue.edu](mailto:rajkumav@purdue.edu)

N. K. Pulipeta · B. G. Musalaya  
University of Central Missouri, Warrensburg, MO, USA  
e-mail: [nxp79750@ucmo.edu](mailto:nxp79750@ucmo.edu)

B. G. Musalaya  
e-mail: [bxm09770@ucmo.edu](mailto:bxm09770@ucmo.edu)

## 1 Introduction

Joint classification models are a promising approach for predicting multiple output variables in machine learning. Unlike traditional binary classification models, they can estimate the probabilities of multiple output variables. This can be particularly useful in natural language processing, where joint models can handle multiple tasks simultaneously or sequentially. These models use shared linguistic features to enhance performance and generalization capabilities.

Correctly modeling joint classification models can lead to significant improvements in both accuracy and training time. However, careful consideration must be given to various factors such as model architecture, the selection of appropriate attributes to model together, and choosing the right learning representation. If these factors are not selected correctly, the performance of joint models may deteriorate instead of improving. Therefore, it is essential to tune and optimize these factors to build accurate and efficient joint models. By doing so, we can leverage the shared features across multiple tasks and improve the performance of our models, leading to better efficiency and use of data.

In this work, we focus on the use of joint models for predicting political bias and other attributes such as topic and source in news data. We study and demonstrate the importance of selecting the appropriate attributes to model together, the model architecture, and the learning representation for improving the performance of bias prediction. Specifically, we conduct experiments on a political bias dataset to investigate the impact of these factors on joint modeling performance. Our results highlight the importance of carefully selecting attributes and designing models for improving the accuracy and efficiency of joint models for news data prediction.

## 2 Related Work

In this work [1] for the news articles, the authors have predicted the political ideologies and also proposed an adversarial media adaptation approach. Authors [2] have proposed a framework that detects and identifies the biases in the text data and for the detection they used BERT [3]. In news articles for automatic detection of bias, the authors [4] have proposed a headline attention mechanism [5]. The authors propose a methodology that quantifies political bias using various word embeddings; they have adapted techniques from gender bias research. They have taken a large corpus of tweets from US politicians, revealing trends, and discussed the challenges of modeling biases along multiple dimensions. Even they have used the nlp techniques for healthcare [6, 7] domain too.

### 3 Problem Statement

In the context of the joint modeling for political bias dataset, we answer the following research questions. Does joint modeling for a political bias dataset improve the performance in terms of accuracy and time? If yes, is there any importance of relation between two attributes that are learnt together? How do different representation learning and model architecture affect the performance improvement?

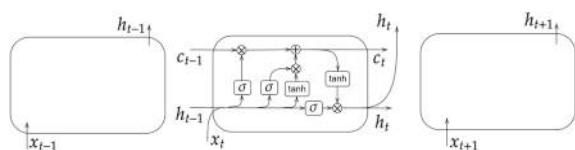
### 4 Background on NLP Methods

In natural language processing (NLP), representation learning entails turning text into numerical vectors that represent syntactic and semantic information. This procedure enables machines to efficiently comprehend and handle human language. Words, phrases, and sentences are effectively represented in terms of their meaning and context. Many NLP tasks, including sentiment analysis, text categorization, and machine translation, are based on these representations. NLP models can develop better representations, leading to higher prediction accuracy and improved performance. We use LSTM and BERT models to learn the representation of the sentence. The following sub-sections give a brief introduction on LSTM and BERT architectures.

#### 4.1 LSTM

An LSTM neural network, which is a variation of RNN architecture, has a unique forget gate that determines whether or not to move information from one memory cell to the next. Figure 1 shows the structure of a single LSTM cell. The sigmoid activation function ( $\sigma$ ) placed on the left side functions as a forget gate and represents the  $t$ th cell in the figure's central panel. The rate at which cells exchange information in the form of memory is indicated by the scalar  $c$ . This implies that, should  $\sigma$  output zero, it will erase the preceding memory  $c_{t-1}$  through multiplication by zero, but if  $\sigma$  outputs one, it will retain values for  $c_{t-1}$  for use in the following cell. It calculates an input value for memory updates using the middle sigmoid activation function, called an input gate. On the other hand, there exists a rightmost sigmoid function

**Fig. 1** Schematic illustration of a memory cell in an LSTM neural network.  
Source [8]



which acts as an output gate that determines what  $h_t$  turns out to be. Additionally, the hyperbolic tangent function produces the output  $h_t$ .

### (1) Forget Gate

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

### (2) Input Gate

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

### (3) Cell State

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

### (4) Output Gate

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

### (5) Hidden State

$$h_t = o_t * \tanh(C_t) \quad (6)$$

- $\sigma$  is the sigmoid activation function.
- $\tanh$  is the hyperbolic tangent function.
- $W_f, W_i, W_C, W_o$  are the weight matrices for the forget gate, input gate, candidate memory cell, and output gate, respectively.
- $b_f, b_i, b_C, b_o$  are the bias vectors for the forget gate, input gate, candidate memory cell, and output gate, respectively.
- $\odot$  denotes element-wise multiplication.
- $[h_{t-1}, x_t]$  denotes the concatenation of the previous hidden state  $h_{t-1}$  and the current input  $x_t$ .

## 4.2 BERT

Google came up with a more advanced NLP model named Bidirectional Encoder Representations from Transformers (BERT). It has a transformer architecture unlike

the traditional models, which is designed for bidirectional training where the context of words is taken into account both from right and left side. This makes BERT be able to understand language in greater depth due to its ability to look at words in surrounding sentences or paragraphs. Multiple layers of encoders each having self-attention mechanisms make it possible for the model to assign significance dynamically.

## 5 Methodology

The key problem in any textual classification is learning the representation of the sentence, which significantly affects the problem. In our case, we are attempting to learn a representation that can aid in joint classification. As a part of this work, we experimented with LSTM (which views each news article as a sequence of tokens) and BERT (which uses an attention mechanism for structured learning) to contrast different types of representation learning approaches. More details on the different architectures and attributes on which we ran experiments are explained in the Experiments section.

We faced two challenges in applying these methods. The first challenge was with the availability of the dataset. Initially, we wanted to experiment with political bias—sentiment-based joint classification. However, since none of the publicly available datasets had sentiment data tagged, we came up with a hybrid pseudo-labeling approach to label the news data with sentiment tags. The details of this approach are mentioned in the dataset preparation section.

To verify the validity of our methods, we wanted to run the experiments multiple times and calculate the average accuracy increase. However, since BERT and LSTM were computation-intensive tasks, we could only test the applied methods on a sub-sample of the dataset, which had a uniform distribution among all the classes.

## 6 Dataset

### 6.1 *Dataset Description*

The dataset selected for the experiments was sourced from a well-known study [1], which collected the data from Allsides (<http://allsides.com/>). The dataset has a cardinality of 34,737 data samples and is publicly available. The dataset is pre-augmented with many interesting attributes such as topic, bias, source, and cleaned text (for NLP experiments). The pre-augmented attributes (topic and source) are used in our experiments to answer the aforementioned research questions.

## 6.2 Dataset Preparation

The dataset consists of 34,737 articles published by 73 news media outlets and covering 109 topics. To enhance the efficiency of our analysis and to address the long-range dependency problem inherent in LSTM models, we selected a subset of 4070 data samples where each article's length was less than 512 tokens. This selection allows us to more effectively compare the performance of LSTM and BERT models. After subsampling the data, our dataset covered data from 105 topics and 60 news media outlets.

## 6.3 Text Preprocessing

Before analyzing the text data, we applied preprocessing methods aimed at making it more contextual and reducing its dimensions. Initially, we used Natural Language Toolkit (NLTK) stop words list to get rid of stop words. As a result, we managed to eliminate such common though useless words as “and,” “the,” “is” and alike that do not add much sense to what is being said. Afterward, we did lemmatization so as to change the word form to its base or dictionary form. For example, words like running, ran, and runs got converted into run their lemma. In this case, lemmatization was important because it reduces variations in words while keeping their meaning intact.

By systematically removing stop words, lemmatizing, and stemming, we effectively normalized the text data, making it more suitable for subsequent analysis and modeling.

## 6.4 Pseudo-Labeling of Sentiment Data

In all the datasets we explored, there were no sentiment labels available. Therefore, we used a form of pseudo-labeling based on the predictions of three state-of-the-art models: XLNet [10], RoBERTa [11], and XLM-RoBERTa [12], which were trained on different datasets for sentiment analysis. This approach allowed us to generate approximate sentiment labels through majority voting.

### 6.4.1 Hybrid Pseudo-Labeling Approach

For pseudo-labeling, three trained models (SOTA models for sentiment prediction) are used to predict the sentiment of each unlabeled article, and the majority vote of the three models predictions is used to determine the final sentiment label for each article. The sentiment label is then assigned to the corresponding article in the

**Table 1** Attribute and number of classes

Attribute	Number of classes
Bias	3
Topic	105
Sentiment	2
Source	149

dataset. The labeled sentiment data is combined with the original labeled data to create a new labeled dataset.

Since the sentiment labels from each of XLNet, RoBERTa, and XLM-RoBERTa will not be accurate, we do majority voting by combining the predictions of these models to improve the overall performance and reliability of the final output as shown in Fig. 1. Majority voting is employed to enhance the accuracy, robustness, and reliability of the final output by leveraging the strengths of multiple models, capitalizing on their diversity and complementarity, and reducing the impact of individual errors or biases.

## 6.5 Dataset Statistics

As mentioned, we took 4070 samples out of the 34,737 articles present, and the size of each of the article selected is less than or equal to 512 token. The attributes and the number of classes are provided in Table 1.

# 7 Experiments

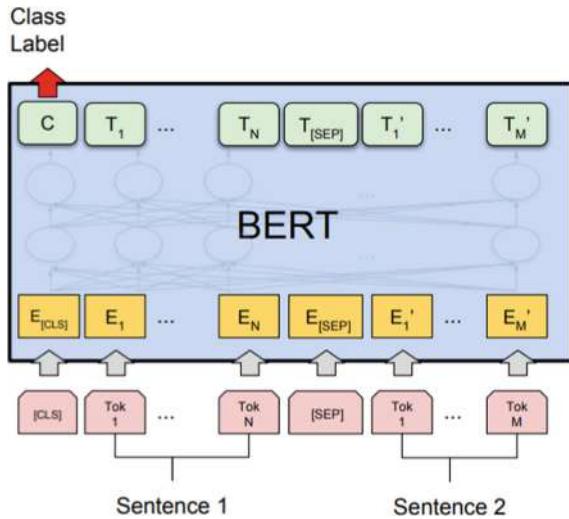
This section outlines the various experiments conducted in predicting bias jointly with other attributes. The following three sub-sections provide information on the different variations of model architecture for joint learning, representation learning, and joint attributes, respectively.

## 7.1 Model Architectures

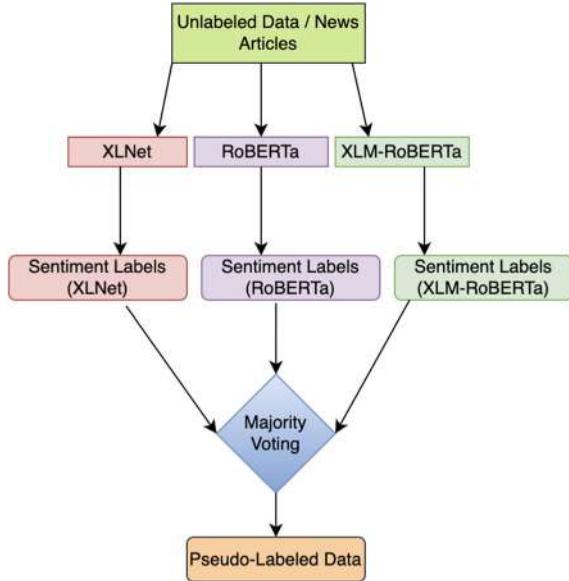
We tried two different architectures for joint modeling, described in Figs. 2 and 3.

- In the first architecture (Fig. 2), the two classes that are modeled together are combined into one fully connected layer. If the number of class labels in each of

**Fig. 2** BERT architecture for sentence pair classification. *Source* [9]



**Fig. 3** Pseudo-labeling of sentiment data



the attributes is  $m$  and  $n$ , the architecture has  $m * n$  outputs. Each of the outputs corresponds to a combination of classes from both attributes.

- In the second architecture (Fig. 3), the two classes that are modeled together are represented by using separate fully connected layers. If the number of class labels in each of the attributes is  $m$  and  $n$ , the architecture has  $m+n$  outputs.

## 7.2 Representations

We have experimented with two different representations—LSTM and BERT. For LSTM-based approach, GloVe embeddings were used and BERT tokenizer was directly used for BERT-based approach.

## 7.3 Attribute Selection

We tried modeling bias with three different attributes—sentiment, bias, and source.

## 8 Results

The section describes the results presented in Figs. 5, 6, and 7. For each joint modeling result introduced below, all experiments and corresponding results are included in the appendix section.

Regarding joint modeling for bias-sentiment, we conducted experiments using both LSTM and BERT for representation learning and utilized the two model architectures described in Sect. 4.1. Figure 4 presents a comparison of the joint modeling results of our best model, selected based on hyperparameter tuning, to a baseline model trained using LSTM/BERT with a single fully connected layer at the end.

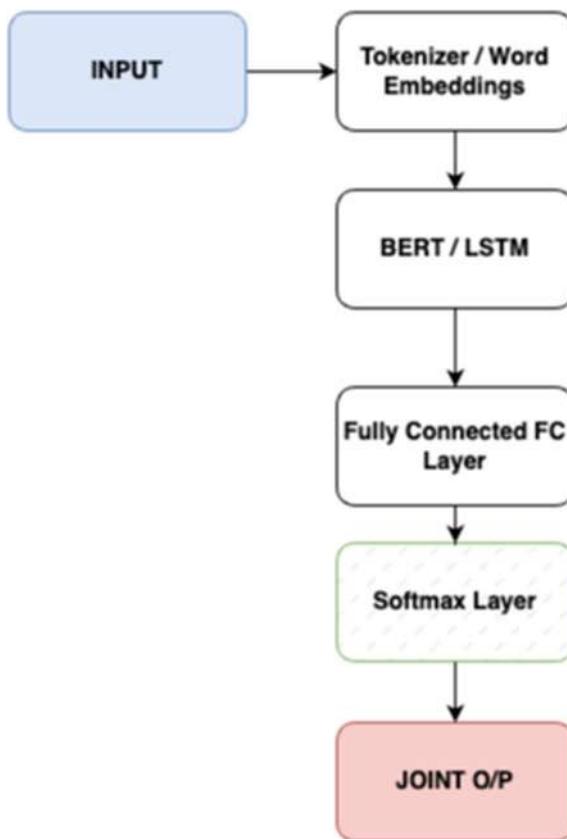
Regarding joint modeling for bias-topic and bias-source, we conducted experiments only using both LSTM for representation learning and utilized the two model architectures described in Sect. 4.1. Figures 5 and 6 presents a comparison of the joint modeling results of our best model for respective joint modeling, selected based on hyperparameter tuning, to a baseline model trained using LSTM with a single fully connected layer at the end (Tables 2, 3, 4, 5, 6, 7, 8, 9 and 10).

## 9 Analysis

This section briefly describes the findings and observations made based on the results obtained across the set of experiments and also answers the research questions raised in the research questions section.

1. Does joint modeling for a political bias dataset improve the performance in terms of accuracy and time?

From Figs. 5, 6, and 7, it can be clearly observed that joint modeling has boosted the performance of bias prediction across all joint modeling combinations of separate fully connected layer based models and for the joint fully connected



**Fig. 4** Joint FC layers

layer for bias-sentiment prediction. In terms of time, we have noted a speedup of 1.7x when modeled jointly, in contrast to training the baselines separately.

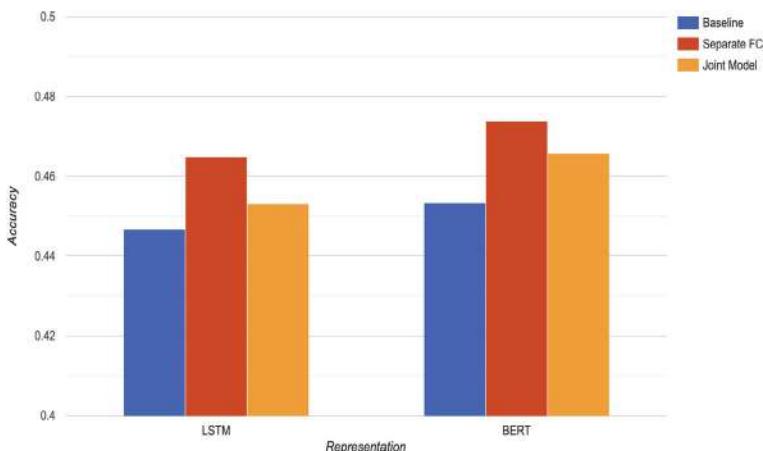
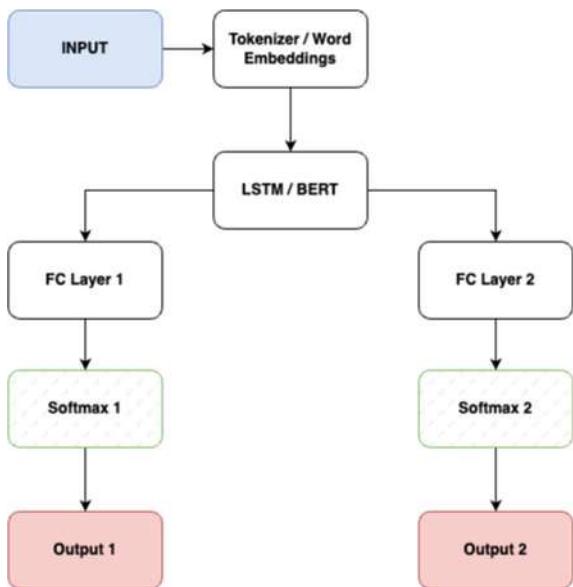
2. If yes, Is there any importance of relation between two attributes that are learnt together?

Attribute selection greatly affects the performance of the model. When we performed CHI-Square test, we have observed that the co-relation between bias and sentiment/topic/source is in the order of source, topic, and sentiment in descending order.

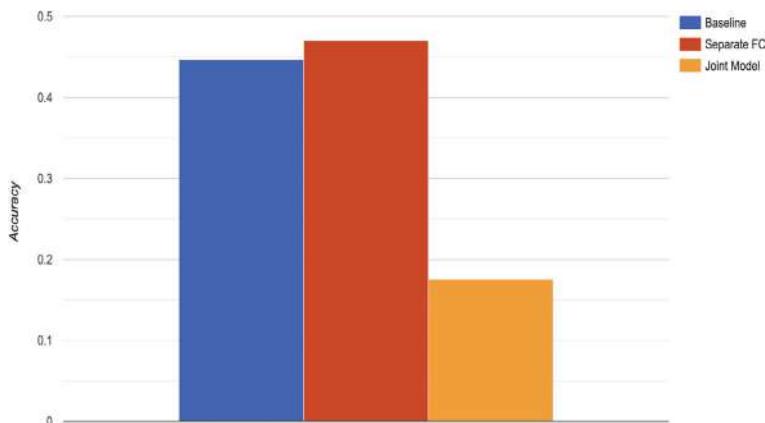
Highly correlated attributes provide a better performance boost than low correlation attributes. This can be attributed to direction of loss, which is boosted better when we train highly correlated attributes.

3. How does different representation learning and model architecture affect the performance improvement?

From Figs. 5, 6, and 7, we can see that modeling the attributes as separate FC layers boosts the performance of bias prediction as the accuracy values are higher

**Fig. 5** Separate FC layers**Fig. 6** Performance of bias prediction when modeled with sentiment

than the baseline accuracy. However, modeling them using a combined FC layer, represented as “joint model” in the results, gives a slight boost if modeled with sentiment and reduce the performance for other two attributes. This could be attributed to the number of classes in the joint labels. Number of labels is much lesser in the sentiment-bias combination when compared to topic-bias and source-bias combinations and the dataset is too small to learn 150+ output



**Fig. 7** Performance of bias prediction when modeled with topic

**Table 2** Baseline performance of LSTM

Hidden nodes	Learning rate	Bias accuracy	Loss
25	1.00E-03	0.4466	1.1049
50	1.00E-03	0.4226	1.1289
100	1.00E-03	0.4201	1.0836
200	1.00E-03	0.3963	1.1050
25	1.00E-05	0.4405	1.0744
25	1.00E-04	0.4467	1.0575
25	1.00E-03	0.4160	1.0742
25	1.00E-02	0.4437	1.1077
25	1.00E-01	0.4437	1.1077

**Table 3** Baseline performance of BERT

Hidden nodes	Learning rate	Bias accuracy	Loss
25	1.00E-03	0.3133	1.2382
50	1.00E-03	0.3133	1.2382
100	1.00E-03	0.3133	1.2382
200	1.00E-03	0.3245	1.2453
300	1.00E-03	0.4533	1.0981

labels. With a larger dataset, the joint architecture could behavior differently (Fig. 8).

**Table 4** Performance of BERT having separate FC layers for bias and sentiment

Hidden nodes	Learning rate	Bias accuracy	Sentiment accuracy	Loss
25	1.00E-03	0.4533	0.6462	3.1656
50	1.00E-03	0.4453	0.6437	3.1564
100	1.00E-03	0.4674	0.6412	3.1972
200	1.00E-03	0.4739	0.6566	3.1810
300	1.00E-03	0.4437	0.6491	3.1723

**Table 5** Performance of LSTM having joint FC layer for bias and source

Hidden nodes	Learning rate	Joint accuracy	Loss
50	1.00E-05	0.0993	5.0023
50	1.00E-04	0.1145	4.9829
50	1.00E-03	0.1686	4.8476
100	1.00E-04	0.1722	4.9259
100	1.00E-03	0.1686	4.8469
200	1.00E-04	0.1686	4.8609
200	1.00E-03	0.1686	4.8469

**Table 6** Performance of LSTM having separate FC layers for bias and sentiment

Hidden nodes	Learning rate	Bias accuracy	Sentiment accuracy	Loss
25	1.00E-03	0.4466	0.6505	3.1681
50	1.00E-03	0.4466	0.6505	3.1681
100	1.00E-03	0.3889	0.6553	3.1386
200	1.00E-03	0.4157	0.6491	3.1829
50	1.00E-05	0.4376	0.6489	3.1078
50	1.00E-04	0.4649	0.6899	3.0544
50	1.00E-03	0.4450	0.6491	3.1301
50	1.00E-02	0.4437	0.6491	3.1723
50	1.00E-01	0.4437	0.6491	3.1723

## 10 Conclusion and Future Work

The observations made from the experiments conducted as a part of this work have yielded the following inferences that training joint modeling could improve performance in terms of both accuracy and training time. It is very interesting to observe that having a separate fully connected layer-based architecture has consistently better performance for joint modeling tasks even when cardinality of the number of classes

**Table 7** Performance of LSTM having joint FC layer for bias and sentiment

Hidden nodes	Learning rate	Joint accuracy	Loss
25	1.00E-03	0.2930	1.7506
50	1.00E-03	0.2930	1.7506
100	1.00E-03	0.2779	1.7490
200	1.00E-03	0.2904	1.7503
50	1.00E-05	0.2899	1.7527
50	1.00E-04	0.2953	1.7239
50	1.00E-03	0.2764	1.7301
50	1.00E-02	0.2904	1.7532
50	1.00E-01	0.2904	1.7532

**Table 8** Performance of BERT having joint FC layer for bias and sentiment

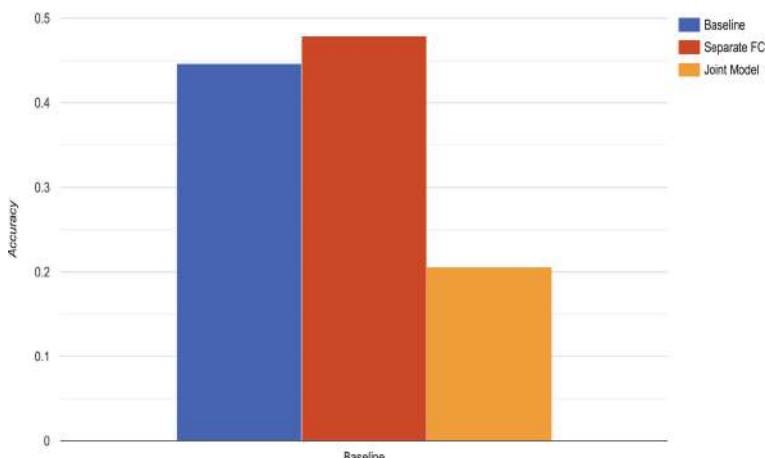
Hidden nodes	Learning rate	Joint accuracy	Loss
25	1.00E-03	0.1941	1.8495
50	1.00E-03	0.1941	1.8495
100	1.00E-03	0.3133	1.2382
200	1.00E-03	0.1941	1.8495
100	1.00E-05	0.2998	1.7327
100	1.00E-04	0.2998	1.7438
100	1.00E-03	0.1941	1.8495
100	1.00E-02	0.2998	1.7438
100	1.00E-01	0.1941	1.8495

**Table 9** Performance of LSTM having separate FC layers for bias and topic

Hidden nodes	Learning rate	Bias accuracy	Topic accuracy	Loss
50	1.00E-05	0.4386	0.0204	9.0222
50	1.00E-04	0.4636	0.0521	8.9691
50	1.00E-03	0.4437	0.1273	8.8573
100	1.00E-05	0.4435	0.0381	8.9945
100	1.00E-04	0.4700	0.0830	8.9452
100	1.00E-03	0.4437	0.1273	8.8559
200	1.00E-05	0.4437	0.0378	8.9698
200	1.00E-04	0.4690	0.1273	8.9005
200	1.00E-03	0.4437	0.1273	8.8559
300	1.00E-05	0.4437	0.0835	8.9609
300	1.00E-04	0.4437	0.1273	8.8676
300	1.00E-03	0.4437	0.1273	8.8559

**Table 10** Performance of LSTM having joint FC layer for bias and topic

Hidden nodes	Learning rate	Joint accuracy	Loss
50	1.00E-05	0.0307	5.6729
50	1.00E-04	0.0518	5.6669
50	1.00E-03	0.0600	5.6216
100	1.00E-05	0.0516	5.6724
100	1.00E-04	0.0555	5.6555
100	1.00E-03	0.0600	5.6197
200	1.00E-05	0.0563	5.6694
200	1.00E-04	0.0597	5.6276
200	1.00E-03	0.0600	5.6193
300	1.00E-04	0.0582	5.6231
300	1.00E-03	0.0543	5.6244

**Fig. 8** Performance of bias prediction when modeled with source

increased and the performance boost is directly proportional to correlation between two attributes being learnt along. Though correlation also played some key in improving the performance for single layer-based joint prediction, the performance started dropping as the number of classes started to increase.

## References

1. Baly R, Martino GD, Glass J, Nakov P (2020) We can detect your bias: predicting the political ideology of news articles. In: Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP), pp 4982–4991

2. Raza S, Garg M, Reji DJ, Bashir SR, Ding C (2024) Nbias: a natural language processing framework for bias identification in text. *Expert Syst Appl* 237:121542
3. Devlin J, Chang MW, Lee K, Toutanova K (2018) Bert: pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*
4. Gangula RR, Duggenpudi SR, Mamidi R (2019) Detecting political bias in news articles using headline attention. In: Proceedings of the 2019 ACL workshop BlackboxNLP: analyzing and interpreting neural networks for NLP, pp 77–84
5. Gordon J, Babaeianjelodar M, Matthews J (2020) Studying political bias via word embeddings. Companion proceedings of the web conference 2020:760–764
6. Bondugula RK, Udgata SK, Rahman N, Sivangi KB (2022) Intelligent analysis of multimedia healthcare data using natural language processing and deep-learning techniques. In: Edge-of-things in personalized healthcare support systems. Elsevier, pp 335–358
7. Bondugula RK, Udgata SK (2023) Novel deep learning models for optimizing human activity recognition using wearable sensors: an analysis of photoplethysmography and accelerometer signals. In: International conference on frontiers of intelligent computing: theory and applications. Springer, pp 45–56
8. Chao Z, Molitor D, Needell D, Porter MA (2022) Inference of media bias and content quality using natural-language processing. *arXiv preprint arXiv:2212.00237*
9. Lee K, Toutanova K, Devlin J, Chang M-W (2018) Bert: pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805v1*
10. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov RR, Le QV (2019a) Xlnet: generalized autoregressive pretraining for language understanding. *Adv Neural Inf Process Syst* 32
11. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V (2019a) Roberta: a robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*
12. Conneau A, Khandelwal K, Goyal N, Chaudhary V, Wenzek G, Guzmán F, Grave E, Ott M, Zettlemoyer L, Stoyanov V (2019) Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*

# Sentiment Analysis and Assessment of Public Opinions Regarding COVID-19 Vaccination via Twitter and Machine Learning Techniques



Roa'a Mohammedqasem, Hayder Mohammedqasim, Bilal A. Ozturk,  
Layth Mhmod Farhan, and Abualqasim Khalil Ismael

**Abstract** Social media has emerged as the most important platform for individuals, organizations, and governments to communicate their views in the world. The same proves to be crucial that social media sites have played during the pandemic of coronavirus disease 2019 (COVID-19), wherein people communicated, shared, and expressed what they perceived. Additionally, betterment of response to such loads pertaining to time-critical issues can be done through analysis of such textual data. Text data preprocessing includes cleaning, tokenizing the text, and stemming followed by the conversion of text data into numerical vectors through CountVectorizer. The Logistic Regression and SVC algorithms were used. The data is then fitted into the SGDClassifier model and further prepared for evaluation against accuracy scores and classification reports. The best result was achieved with the SGDClassifier algorithm with CountVectorizer, with an accuracy of 88%, a precision of 87.9%, a recall of 87.9%, an F1-score of 87.8%, and AUC of 86%, respectively. The results of the study highlight the significance of class-imbalance management and prove that machine learning techniques are potent to decide the public sentiment at the real-time level. This work will serve as a benchmark for the future in the promising area, and it is in line with how the social media data can be utilized to take into account public sentiments in a pandemic situation.

---

R. Mohammedqasem · H. Mohammedqasim · B. A. Ozturk · L. M. Farhan (✉) · A. K. Ismael  
Faculty of Engineering, Istanbul Aydin University, Istanbul, Turkey  
e-mail: [laythfarhan@stu.aydin.edu.tr](mailto:laythfarhan@stu.aydin.edu.tr)

R. Mohammedqasem  
e-mail: [rmoammedqasem@aydin.edu.tr](mailto:rmoammedqasem@aydin.edu.tr)

H. Mohammedqasim  
e-mail: [hmoammedqasim@aydin.edu.tr](mailto:hmoammedqasim@aydin.edu.tr)

B. A. Ozturk  
e-mail: [bilalo@aydin.edu.tr](mailto:bilalo@aydin.edu.tr)

A. K. Ismael  
e-mail: [akhalilismael@stu.aydin.edu.tr](mailto:akhalilismael@stu.aydin.edu.tr)

**Keywords** Social media · Natural language processing (NLP) · Study of Covid-19 on Twitter · Sentiment analysis · Vaccination · Machine learning

## 1 Introduction

The Covid-19 epidemic, which started in late 2019, has resulted in the confirmation of tens of millions of cases. Infections and a global death toll approaching millions [1]. The epidemic has brought large-scale economic and social upheaval, changing the everyday life of people around the world. Protection the implementation of such measures has proved to be quite effective in slowing the spread, but a concerted worldwide effort went into the creation of Covid vaccines aims to directly confront the illness and achieve immunity. Research indicates that herd immunity requires vaccination of at least 70% of the population [2]. Equally, implementation varied greatly from area to area as circumstances diverged. Furthermore, disputes over civil liberties and doubts about the efficacy of vaccines presented challenges. Nobody is safe until everybody is safe. The pandemic has seen a significant proportion of the population raise questions about the safety of the Covid vaccination has been questioned due to skepticism generated on social media platforms, resulting in worries, hesitance, and opposition toward the distribution procedures of the vaccine [3]. This dataset is utilized to analyze English tweets and investigate overall feelings and opinions pertaining to the Covid-19 epidemic. The tweets were gathered in the early stages of the pandemic, specifically from March 16, 2020.

Since public perceptions and public endorsement of the Covid-19 vaccine are applicable to the ability to vaccinate major population groups to reach herd immunity, these constitute important researcher considerations. Although many reports have been made on the financial and social repercussions of Covid [4] and dissemination dynamics, once again, little has been reported about the problem. In the digital era, Twitter has emerged as a pivotal platform for public discourse, reflecting widespread opinions on global events, including public health emergencies such as the COVID-19 pandemic. The analysis of sentiments expressed in tweets provides critical insights that can inform government policies, public health strategies, and community responses. The scope of this research centers on enhancing the accuracy and efficiency of sentiment analysis techniques applied to Twitter data, focusing specifically on discussions related to COVID-19 vaccines [2].

The major aim of this research is the public discourses taking place through social media, particularly the platforms of Twitter, against the ongoing drives of vaccination being done across the globe against Covid-19. Currently, the need for any policymakers and health organizations is to understand public sentiment and public reaction toward the rollout of the Covid-19 vaccine. In contrast, there is little research about the implications of sentiment analysis using Twitter data on public opinion and sentiments regarding the Covid-19 vaccine. There is always natural noise and variability in social media data, and it would be impractical to aim for the same level of accuracy as might be possible with other types of data. This article

will rather focus on the robustness and the reliability of the insights, even if some sort of necessary elimination in the accuracy has to be made to obtain results that are valuable for the purpose of decision-making. Here, it is important that these assessments analyze the limitations of machine learning models to capture human sentiment completely, explaining them in that light with respect to the results.

This paper considers the ensemble technique, using natural language processing and machine learning approaches to build a model that can predict the accuracy of contributes to sentiment analysis in public health communication, especially during the Covid-19 vaccination campaign. The proposed model, therefore, uses the traditional machine learning models to measure the achieved performance: AUC, accuracy, recall, accuracy, precision, and F1-score. Relied on the feature extraction of CountVectorizer: In general, the study makes the following main contributions applicable in a strong model with improved accuracy:

- NLP Techniques Used: Utilized natural language processing techniques to process and analyze text data.
- Application of the Coronavirus tweets: The testing data is from Kaggle, and it is part of the training data for the purpose of training the machine learning models.
- Feature extraction with CountVectorizer: Converts a collection of text documents to a count matrix, which contains the token's count. Many machine learning algorithms take this matrix as input.
- Application of machine learning models: Training and evaluation of diverse machine learning models on the data which includes classifiers with the SGDClassifier, Logistic Regression, and SVC algorithms. The training and evaluation of the models is based on the different performance metric of AUC, Recall, Accuracy, Precision, and F1 score.

## 2 Literature Review

The Covid-19 virus is briefly discussed. Sentiment analysis was performed on Twitter data regarding the closure in India during the COVID-19 pandemic by Gupta et al. [5] the data collection comprises 12,741 tweets specifically related to the keyword “Indialockdown” sourced from Twitter. The process of assigning labels to the data is carried out via the VADER and TextBlob lexicons. They have employed a total of eight. The study employs metrics such as precision, recall, accuracy, and F1-score. Through their experiments, it was found that the LinearSVC algorithm used in comparison with other models, the unigram model demonstrated the maximum accuracy of 84.4%. Jalil et al. [6] this research made use of the Twitter data that was collected to establish the public sentiment toward COVID-19 from February to March 2020. A higher-level multi-depth DistilBERT model using a Transformation-based technique analyzed the sentiment of the research. The classification accuracy of the technique was below 85%, which means there is a challenge in the effective classification of sentiments toward noisy data in social media. Jelodar et al. [7] utilized a dataset consisting of 563,079 Reddit posts for their investigation. Topic

modeling has been implemented through the utilization of LDA. The posts have been categorized into five sentiment scores, which range from extremely positive to highly negative. The LSTM model is contrasted with a variety of machine learning algorithms, including Naïve Bayes, SVM, KNN, and logistic regression. The LSTM model demonstrates superior performance, achieving an accuracy of 81.15% which surpasses the accuracy of the other models. Talaat et al. [8, 9] A hybrid BERT-based model was used in this research for sentiment classification of COVID-19 tweets. Although the model used several layers of BiRNN to boost contextual understanding, it gave an accuracy of less than 85%, showing how sentiment classification accuracy remained a challenge in improving with existing methodologies. Parveen et al. [9] this research applied the TextBlob and Naive Bayes algorithms for sentiment classification of COVID-19-related tweets, from which an accuracy of about 82% was achieved, which speaks to the difficulty of correctly managing variability and brevity in Twitter data. Li et al. [10] analyzed a dataset of 3000 Weibo postings, which were categorized manually into seven distinct groups. The prediction task utilized SVM, Naive Bayes, and random forest classifiers. Among these, the random forest classifier demonstrated superior performance compared to the others, obtaining an accuracy rate of 81%. Tan et al. [11] this research paper analyzed public sentiment toward COVID-19 vaccines using machine learning classifiers and preprocessing and feature-extraction methods. The expected performance for this best-fitting model achieved an 83% accuracy of the sentiments from varied and often ambiguous data on social media. Pristiyono et al. [12] conducted an analysis of tweets from January 15 to 22, 2021. They employed hand labeling and WordCloud for feature extraction, and Naïve Bayes for classification. The study revealed a significant prevalence of negative emotion, amounting to 56%, and an accuracy rate below 84%. They have deduced from the WordCloud. Nemes and Kiss [13] conducted sentiment analysis on tweets, posts, hashtags, and comments using Recurrent Neural Networks (RNN) for categorization. The RNN model was compared to TextBlob and surpassed it, obtaining an accuracy of less than 82.2%. The suggested recurrent neural network (RNN) model is compared to TextBlob, and it demonstrates superior performance over the latter model.

### 3 Methodology

The data collection for Covid-19 Twitter data using a purposeful sampling strategy. The data analysis is categorized into two main sections. The initial section delves into an investigative examination of tweets, data visualizations, and a depiction of the fundamental attributes of Covid-19 vaccination Twitter data. The objective of this technique is to provide understanding and analysis of public responses and conversations on Twitter about the global implementation of Covid-19 immunization campaigns. The second section focuses on the sentiment categorization of tweets using supervised machine learning methods. We opted for a supervised machine learning methodology due to the presence of accurate and comprehensive labels in

our dataset. Supervised learning enables us to assess the accuracy ratings of the selected classifiers during sentiment classification.

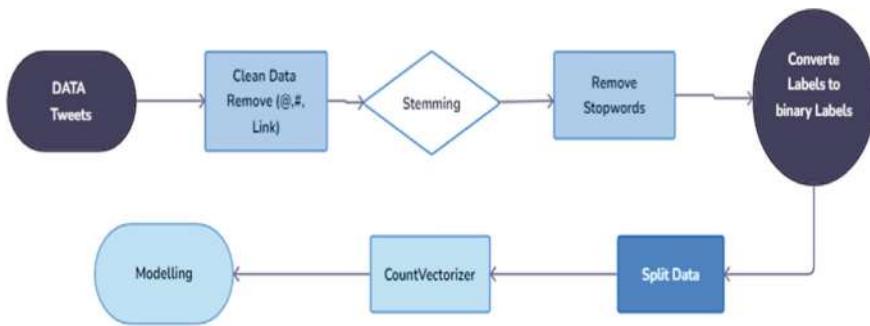
### ***3.1 Data Collection***

Among other sources, the dataset was collected from Kaggle, which is one of the platforms hosting a huge repository of datasets and computational competitions. In this work, the dataset will be analyzed in English tweets, concerning overall attitudes and opinions regarding the Covid-19 epidemic. Concerning the sentiments made, the tweets were collected in the early stages of the epidemic on March 16, 2020. The dataset attributes are varied and include the location of the user, date of posting tweet, tweet content, and sentiment. Sentiments associated with the tweet are classified into different classes, such as “Neutral,” “Positive,” or “Extremely Negative,” among others [14]. This dataset was collected from the Twitter API, and as such, it provides a wide aggregation of tweets specifically related to COVID-19 vaccinations. The aim was to understand public perceptions online in terms of the vaccination trends of the pandemic.

### ***3.2 Data Preparation***

It ensures that there is proper cleaning and preprocessing of tweets to enhance the quality of the dataset; hence, ensures that the dataset is of good quality. Due to the effect of abnormalities, tweets are usually noisy with the inclusion of punctuation marks, symbols, @links, stop words, and other special characters. A python script does the cleaning of such inessential elements in collected tweets to guarantee that the results are more precise [15]. A python script removes irrelevant or unwanted elements from the collected tweets, such as spaces, punctuation, hashtags, URLs, special characters, hyperlinks, and common terms. The first step is data preparation, in which the unwanted or irrelevant text elements are removed [16]. The remove\_pattern is a method to search for the specified pattern and replace it with an empty string. On cleaning the text data, all the characters are cleaned except for the letters and the hashtag symbol, so the analysis will only be on the text content and hashtags, which may bear a sentiment value. The tweets were processed by converting them to a consistent case format, removing short words with less than three characters, and tokenizing them for Natural Language Processing (NLP) techniques like stemming. This normalizes text data by reducing inflected words to their stem, base, or root form. Stopwords were removed from the tweets using the NLTK library, as they contribute little to sentiment understanding. This ensures focus remains on more meaningful words, as shown in Fig. 1.

Data labeling is the process of representing each data point with its descriptive tags or category. Objective of the Study: This study aims at categorizing the tweets



**Fig. 1** Flowchart for the proposed model

as done with good or negative using human annotation. The final dataset has in total 41,157 distinct tweets where 25,759 tweets are marked as positive and 15,398 are negative. The dataset used for the purpose of the study consists of two columns, they are and labels. In this column, the textual data is featured with the label that it is associated with. This data is then transformed into features which the model uses in its training and prediction in both cases. Data in the “label” column have binary values, denoted as 1 or 0, to point the feelings of the tweet being labeled. The dataset was modified according to the requirements of sentiment analysis by changing the categorization methodology accordingly [17]. The tweets were first categorized into the following five categories: positive, highly positive, negative, highly negative, and neutral. This five-class system was then converted into binary classification to make that easy and focus on the overall sentiment trend. That is, all positive and highly positive tweets were clubbed and labeled as 1 (Positive), and all the negative, extremely negative, and the neutral were clubbed and labeled as 0 (Non-positive).as shown in Fig. 2. Several metadata columns are associated with every tweet in the dataset, but for this analysis, we use only the text content of the tweets. Other columns, such as user information, timestamps, and geo-tags are not included in the analysis to keep the analysis anonymous and within the bounds dictated by data handling rules. The resultant data set contained only the tweet texts and their corresponding new binary sentiment labels, ensuring that we work with a focused and relevant data set for sentiment analysis [18].

The final dataset, used for the binary classification of sentiments, was constructed by mapping the original multi-class sentiment labels into a binary format. Positive sentiments (both “Positive” and “Extremely Positive”) were coded as 1, and all other

**Fig. 2** Displays a selection of tweets taken from the dataset collected from Twitter

```

binary_dataset = new_df.copy()
binary_dataset["Sentiment"] = binary_dataset["Sentiment"].replace('Positive',1)
binary_dataset["Sentiment"] = binary_dataset["Sentiment"].replace('Extremely Positive',1)
binary_dataset["Sentiment"] = binary_dataset["Sentiment"].replace('Neutral',1)
binary_dataset["Sentiment"] = binary_dataset["Sentiment"].replace('Negative',0)
binary_dataset["Sentiment"] = binary_dataset["Sentiment"].replace('Extremely Negative',0)
  
```

sentiments (including “Neutral,” “Negative,” and “Extremely Negative”) were coded as 0. This binary dataset, consisting of cleaned, stemmed tweets devoid of stopwords and their corresponding binary sentiment labels, forms the backbone of the sentiment analysis model discussed in subsequent sections of this study.

### ***3.3 Feature Extraction***

Text data in an abstract form needs to be numerically converted into features for the machine learning to act upon it. We converted tweet text data into a matrix of token counts by using the CountVectorizer from the sklearn.feature\_extraction.text. At least, that would be one of the most obligatory steps in sentiment analysis, where textual data is molded into a structured format and can further be used by machine learning models. The step will take lexical features and convert them to a feature set that the classifier can use efficiently. One of the performance determinants of the text classification model is feature extraction. We then used the CountVectorizer for feature extraction from the Twitter dataset. The classifiers were then trained with this feature, such that it could capture context information using unigrams and bigrams. Amazingly, the CountVectorizer was reported to have high accuracy back when it used to be part of this sentiment analysis task. A trained classifier that uses features from the CountVectorizer can make accurate sentiment predictions on the test data [19].

### ***3.4 Machine Learning for Sentiment Categorization***

Following the extraction of features, the subsequent activity is to input the feature vectors into the machine learning classifiers for the purpose of conducting sentiment classification. The vaccination tweets were categorized using Stochastic Gradient Descent (SGD) [20], Logistic Regression (LR) [21], and Support Vector Machine (SVM) [22] machine learning models, and their performances were evaluated and compared. The Scikit-learn python library, which is an open-source tool for machine learning, was utilized to access categorization methods for machine learning [23]. During each experiment, the training set is employed to optimize and train the machine learning algorithms, whereas the test set is employed to evaluate the models’ performance [24].

### 3.5 Model Training and Evaluation

The modeling phase of our sentiment analysis study focused on the application and evaluation of several machine learning models to classify tweets into binary sentiment categories based on the prepared textual data. This section describes the model training, validation, and performance evaluation processes. We began by partitioning. This dataset is partitioned into training and validation sets to ease model training and evaluate their ability to generalize on unknown data. Employing the `train_test_split` [25] function from the `sklearn.model_selection` module was implemented. setting aside 20% of the data as the validation set. Stratification was performed to obtain the proportion of emotion labels in the training and validation set, thus representing the overall dataset [26].

## 4 Results

The sentiment analysis findings are obtained by applying three machine learning models to classify COVID-19 vaccine tweets related into binary categories. The evaluation of each model was conducted by assessing its accuracy in categorizing both the training data, which indicates how well the model learnt from the dataset, and the validation data, which indicates how well the model can apply its knowledge to fresh, unknown data.

### 4.1 Main Outcome

Here, we apply two different preprocessing techniques in order to compare performance in two different models. The results differ. Yet, in Table 1, from where CountVectorizer is used, the model Support Vector Machine has an F1, precision, recall—all 0.83, 0.84, 0.83. Logistic Regression is a bit more effective than Support Vector Machine: with an F1, precision, recall, and AUC of 0.86, 0.87, 0.86, 80. Stochastic Gradient Descent turns out to be much better than that: in fact, with the F1, precision, and recall being the same at 0.88. In Table 2, when SMOTE and TF-IDF preprocessing is used, the performance differs. The Support Vector Classifier provided a very balanced performance in the model, showing F1, precision, recall, and AUC of 0.84, 0.85, 0.84, 81, respectively. The Logistic Regression model gives worse performance, with all metrics at 0.82. The performance of SGD also decreased further to an even 0.81. All these results indicate that for both the LR and the SGD models, the CountVectorizer fits better, which comments on the importance of preprocessing methods in the model’s utility. We conducted data analysis on the gathered tweets to ascertain public attitudes, establish keyword linkages, and detect social media patterns pertaining to the implementation. SGD Classifier models are

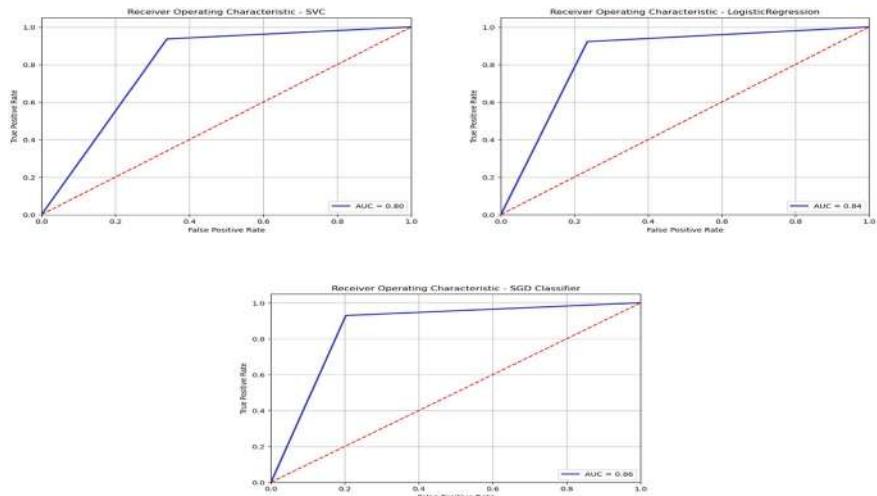
trained on the vectorized data. The performance of these models is evaluated using accuracy scores and classification reports. In the best average accuracy of the model SGDClassifier, the CountVectorizer is 88%. AUC analysis results for the ML models as shown in Fig. 3.

**Table 1** Using CountVectorizer with Twitter COVID-19 dataset

Models	Accuracy (%)	Recall (%)	Precision (%)	F1-score (%)	AUC (%)
SVC	83	83	84	83	80
LR	86	86	87	86	84
SGD	88	87.9	87.9	87.8	86

**Table 2** Using smote and TF-IDF with Twitter COVID-19 dataset

Models	Accuracy (%)	Recall (%)	Precision	F1-Score	AUC (%)
SVC	84	84	0.85	0.84	81
LR	82	82	82%	82%	81
SGD	81	81	81%	81%	80



**Fig. 3** AUC analysis results for the proposed ML models

**Table 3** Comparison to prior works

Study	Method	Accuracy (%)
Gupta et al. [5]	LinearSVC algorithm	84.4
Jelodar et al. [7]	Naïve Bayes, SVM, KNN, and LR	81.15
Talaat et al. [8]	Hybrid BERT-based model and BiRNN	85
Parveen et al. [9]	TextBlob and Naive Bayes algorithms	82
Li et al. [10]	SVM, Naïve Bayes, and random forest classifiers	81
Proposed	SGDClassifier algorithm with CountVectorizer	88

## 4.2 Comparison with Prior Works

This study confirms the promise in the application of machine learning algorithms to evaluate public discussion and opinion in pursuit of the realization of the Covid vaccination distribution campaign. During the research period, a novel trend of Covid-19-related conversations emerged on Twitter, mostly focusing on health education pertaining to the vaccine, its distribution and application, as well as the required dosage for complete immunization. Its availability concerns or questions. These are of evident concern as it directly impacts actual success in getting the vaccination. In Table 3, the trial results confirm the findings in earlier cited studies that have applied information from social media to evaluate the reaction of the public concerning. The n-grams found in this research work are also in line with that in earlier research work carried out on Covid-19, where can check out discussions, opinions on safety, and matters of concern in Twitter.

## 5 Limitation

Due to the nature of social media usage, exposure to sources such as sexist, racist, and other hate speech and derogatory content is present; this makes it controversial in generalizability of the measurement. Other limitations that may apply to an analysis include that it is reliant on social media data, largely based on Twitter, which may not be representative of the population. Besides, noisy variations in the data, temporal changes, and variations in language and culture may further reduce the accuracy of sentiment classification. The current study is based on English tweets, meaning the generalizability toward other languages or cross-cultural settings may be poor. Model generalizability may vary with different datasets or contexts. Advanced models may need retraining or fine-tuning to adapt to the context. Sentiment analysis is further made complex with complex data.

## 6 Conclusion

The objective of our study was to analyze the public dialog and responses on Twitter about the implementation of the Covid-19 vaccine distribution initiative. Well-known machine learning methods were utilized to forecast attitudes based on the gathered tweets. The majority of Twitter users exhibited optimism during the research period; however, a few instances of negative views posed a potential challenge to the overall effectiveness of the vaccination deployment program. Furthermore, we have discovered a novel tendency in discussions relating to Covid-19, which centers around the distribution and implementation of vaccines, the necessary dosage for achieving immunity, and other pertinent health details regarding the vaccine. The models used included Logistic Regression and SVC. Data is then fitted by the SGDClassifier model, which after further prepares for evaluation consisting of the accuracy scores and classification of reports. The best result found was for the SGDClassifier model with CountVectorizer, that gives 88% accuracy, 87.9% precision, 87.9% recall, 87.8% F1-Score, and 86% AUC. These findings can serve as a useful tool for policymakers and relevant authorities to predict and address any issues in the vaccination rollout program. To develop herd immunity against Covid-19, it is crucial to promptly address the concerns of the general population and enhance trust and confidence in the vaccination campaign.

## References

1. Mohammedqasim H, Ahmed Jasim A, Mohammedqasem A, Ata O (2024) Enhancing predictive performance in COVID-19 healthcare datasets: a case study based on hyper Adasyn over-sampling and genetic feature selection. *J Eng Sci Technol* 19(2):598–617
2. Mohammedqasem R et al (2023) Multi-objective deep learning framework for COVID-19 dataset problems. *J King Saud Univ Sci* 35(3):102527. <https://doi.org/10.1016/J.JKSUS.2022.102527>
3. Chawla S, Mehrotra M (2021) Impact of emotions in social media content diffusion. *Informatica* 45(6):11–28. <https://doi.org/10.31449/INF.V45I6.3575>
4. Mohammedqasem R, Mohammedqasim H, Ata O (2022) Real-time data of COVID-19 detection with IoT sensor tracking using artificial neural network. *Comput Electr Eng* 100:107971. <https://doi.org/10.1016/J.COMPELECENG.2022.107971>
5. Gupta P, Kumar S, Suman RR, Kumar V (2021) Sentiment analysis of lockdown in India during COVID-19: a case study on twitter. *IEEE Trans Comput Soc Syst* 8(4):939–949. <https://doi.org/10.1109/TCSS.2020.3042446>
6. Jalil Z et al (2022) COVID-19 related sentiment analysis using state-of-the-art machine learning and deep learning techniques. *Front Public Health* 9:812735. <https://doi.org/10.3389/FPUBH.2021.812735/BIBTEX>
7. Jelodar H, Wang Y, Orji R, Huang S (2020) Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach. *IEEE J Biomed Health Inform* 24(10):2733–2742. <https://doi.org/10.1109/JBHI.2020.3001216>
8. Talaat AS (2023) Sentiment analysis classification system using hybrid BERT models. *J Big Data* 10(1):1–18. <https://doi.org/10.1186/S40537-023-00781-W/TABLES/4>

9. Parveen N, Chakrabarti P, Hung BT, Shaik A (2023) Twitter sentiment analysis using hybrid gated attention recurrent network. *J Big Data* 10(1):1–29. <https://doi.org/10.1186/S40537-023-00726-3/TABLES/10>
10. Li L et al (2020) Characterizing the propagation of situational information in social media during COVID-19 epidemic: a case study on Weibo. *IEEE Trans Comput Soc Syst* 7(2):556–562. <https://doi.org/10.1109/TCSS.2020.2980007>
11. Tan KL, Lee CP, Lim KM (2023) A survey of sentiment analysis: approaches, datasets, and future research. *Appl Sci* 13(7):4550. <https://doi.org/10.3390/APP13074550>
12. Pristiyyono, Ritonga M, Al Ihsan MA, Anjar A, Rambe FH (2021) Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes algorithm. *IOP Conf Ser Mater Sci Eng* 1088(1):012045. <https://doi.org/10.1088/1757-899X/1088/1/012045>
13. Nemes L, Kiss A (2021) Social media sentiment analysis based on COVID-19. *J Inf Telecommun* 5(1):1–15. <https://doi.org/10.1080/24751839.2020.1790793>
14. Banda JM et al (2021) A large-scale COVID-19 Twitter chatter dataset for open scientific research—an international collaboration. *Epidemiologia* 2(3):315–324. <https://doi.org/10.3390/epidemiologia2030024>
15. HaCohen-Kerner Y, Miller D, Yigal Y (2020) The influence of preprocessing on text classification using a bag-of-words representation. *PLoS ONE* 15(5):e0232525. <https://doi.org/10.1371/JOURNAL.PONE.0232525>
16. Imran AS, Daudpota SM, Kastrati Z, Batra R (2020) Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on covid-19 related tweets. *IEEE Access* 8:181074–181090. <https://doi.org/10.1109/ACCESS.2020.3027350>
17. Pak A, Paroubek P. Twitter as a corpus for sentiment analysis and opinion mining. Accessed 27 May 2024. [Online]. Available: <http://tumblr.com>
18. Mohammad SM, Kiritchenko S, Zhu X (2013) NRC-Canada: building the state-of-the-art in sentiment analysis of tweets. \*SEM 2013—2nd joint conference on lexical and computational semantics, vol 2, pp 321–327, Accessed 27 May 2024. [Online]. Available: <https://arxiv.org/abs/1308.6242v1>
19. Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision. Accessed 27 May 2024. [Online]. Available: <http://tinyurl.com/cvvg9a>
20. Vidyashree KP, Rajendra AB (2023) An improvised sentiment analysis model on Twitter data using stochastic gradient descent (SGD) optimization algorithm in stochastic gate neural network (SGNN). *SN Comput Sci* 4(2):1–11. <https://doi.org/10.1007/S42979-022-01607-X/TABLES/5>
21. Qorib M, Oladunni T, Denis M, Ososanya E, Cotae P (2023) Covid-19 vaccine hesitancy: text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset. *Expert Syst Appl* 212:118715. <https://doi.org/10.1016/J.ESWA.2022.118715>
22. Qasim HM, Ata O, Ansari MA, Alomary MN, Alghamdi S, Almehmadi M (2021) Hybrid feature selection framework for the Parkinson imbalanced dataset prediction problem. *Medicina* 57(11):1217. <https://doi.org/10.3390/MEDICINA57111217>
23. Pedregosa F et al (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830. Accessed 27 May 2024. [Online]. Available: <http://scikit-learn.sourceforge.net>
24. Jasim AA, Hazim LR, Mohammedqasim H, Mohammedqasem R, Ata O, Salman OH (2024) e-Diagnostic system for diabetes disease prediction on an IoMT environment-based hyper AdaBoost machine learning model. *J Supercomputing* 1–26. <https://doi.org/10.1007/S11227-024-06082-0/TABLES/4>
25. Mohammedqasim H, Mohammedqasem R, Ata O, Alyasin EI (2022) Diagnosing coronary artery disease on the basis of hard ensemble voting optimization. *Medicina* 58(12):1745. <https://doi.org/10.3390/MEDICINA58121745>
26. Kaur P, Sharma M (2019) Diagnosis of human psychological disorders using supervised learning and nature-inspired computing techniques: a meta-analysis. *J Med Syst* 43(7):1–30. <https://doi.org/10.1007/S10916-019-1341-2/TABLES/14>

# Artificial Intelligence Driven Kyphosis Classification



V. Thamilarasi , R. Harihara Krishnan , V. Vijayalakshmi , J. Mary Catherine , V. Poornima , and S. Pratheepa

**Abstract** This research presented in the paper revolves around the classification of kyphosis disease, a spinal condition characterized by an abnormal curvature of the upper spine, leading to a rounded back. The primary objective is to develop a predictive model that can accurately determine whether a patient has kyphosis based on specific diagnostic measurements. Exploratory Data Analysis was employed to conduct preprocessing tasks with one hot encoding. By investigating machine learning and deep learning algorithms, particularly decision trees and random forest and CNN, this research aims to achieve a high level of accuracy in predicting the presence or absence of kyphosis. Overall, the research contributes to the broader understanding of kyphosis disease classification and highlights the potential of AI-driven techniques in medical diagnostics. Model performance was evaluated by Confusion matrix, precision, Recall, F1 score, and Accuracy. The achieved accuracy level of 93.33% for CNN demonstrates the efficacy of the proposed approach and its relevance in clinical practice for diagnosing and managing kyphosis effectively.

---

V. Thamilarasi ()

Department of Computer Science, Sri Sarada College for Women (Autonomous), Salem, Tamilnadu, India

e-mail: [tamilomsiva@gmail.com](mailto:tamilomsiva@gmail.com)

R. H. Krishnan

Computer Application, Patrician College of Arts and Science, Chennai, India

V. Vijayalakshmi

Department of Computer Science, Christ College of Science and Management, Sonnur, Karnataka, India

e-mail: [vijayalakshmiv@christcollegemalur.com](mailto:vijayalakshmiv@christcollegemalur.com)

J. M. Catherine

Computer Science, Vels Institute of Science, Technology and Advanced Studies, Chennai, Tamilnadu, India

V. Poornima

Chevalier T. Thomas Elizabeth College for Women, Chennai, Tamilnadu, India

S. Pratheepa

Department of Computer Science, J.H.A. Agarsen College, Madhavaram, Chennai, Tamilnadu, India

**Keywords** Kyphosis · Random Forest · Decision tree · CNN · Classification

## 1 Introduction

Kyphosis is a condition where the upper spine curves abnormally, giving the appearance of a rounded back. It must be classified in order to determine the best course of treatment. Important factors to take into account in this classification process are age, number, and underlying causes.

Kyphosis, which can have a number of causes, is characterized by an excessive forward rounding of the upper back. Kyphosis can result from the weakening or cracking of the spinal bones in older people. On the other hand, kyphosis in children or teenagers can result from a malformed spine or gradual wedging of the spinal bones. Though it can occur at any age, kyphosis is typically diagnosed in the 13–16 age range due to rapid bone growth. This may also be influenced by age-related loss of vertebral flexibility. Furthermore, other prenatal medical disorders may coexist with congenital kyphosis. Kyphosis usually doesn't require medical attention, but it can cause issues with one's self-perception. Severe cases may require surgery to treat pain or breathing difficulties.

## 2 Objectives

To identify the suitable algorithm for kyphosis classification.

To identify the correlation among features.

## 3 Literature Review

Hussein et al. [1], provided systematic review of kyphosis prediction and provide quality assessment by STROBE checklist. Author also suggest based on STROBE score all reviewed research papers were quality one and only 2 papers were low quality. This research concluded that AI techniques provide better solution for Kyphosis analysis [1]. Chatter et al. [2], experimented with various machine learning algorithms and suggest ML algorithms provide solution to the kyphosis disease prediction [2].

Chauhan et al. [3], experimented with ML algorithm s RF, SVM, KNN, and DNN with k-fold cross validation and found that RF achieved an accuracy of 83.89%, 85.14% for SVM, 84.03% for KNN, and 87.64% for DNN. As a result, author recommend DNN for kyphosis analysis and it achieved sensitivity of 55%, specificity of 97%, precision of 70%, f1 score of 57% and 76% for AUC-ROC [3]. Dankwa et al. [4], experimented FR, SVM, ANN for kyphosis prediction with 5 and tenfold cross

validation. This research achieved an accuracy of 79–85% for fivefold and 77–86% for tenfold cross validation. As a whole accuracy of ANN 79.12% and 79.03% for 5 and tenfold cross validation which is higher than other models [4].

Passias et al. [5], considered Cervical kyphosis, Cobb angle, Cervical scoliosis, sagittal vertical axis, Cervical sagittal imbalance patient for analysis. Decision Tree was used to experiment and conditional variable table were constructed based on 2000 conditional inferences. Author found Smith-Peterson Osteotomy (OR: 2.55, CI:6.34) were the crucial prediction of DJK [5]. Lafage et al. [6], utilized unsupervised cluster analysis to identify the adult spinal deformity (ASD) and consider 286 patients with hyper-throacic kyphosis, severe sagittal, severe coronal, moderate sagittal clusters. This research recorded that importance of distinct deformity patterns identifications and predict ASD patients surgical outcomes [6].

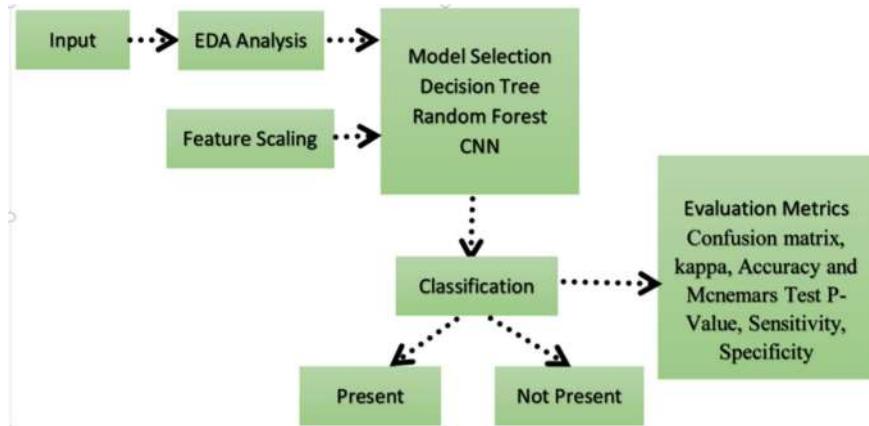
Cina et al. [7], discussed various ML applications such as classification, clustering, and regression for spine care. Author suggest ML algorithms provide better solution in various aspect of spine and suggest this field analysis needs interdisciplinary research and high volume of dataset [7]. Ren et al. [8], analyzed nearly 292 studies for review and articles belongs to image processing, decision supporting, diagnosis, surgery outcomes, rehabilitation, operative assistance, cost, and hospitalization. Author found that the ML algorithm produced better experimental result than other models [8].

Chauhan et al. [9], investigated the machine learning and deep learning models for Kyphosis analysis for various datasets. Author found that deep neural network models achieved an accuracy of 87.72%, Specificity of 0.97%, Precision of 0.76%, F1 Score of 57%, AUC-ROC of 76% and 87.64% for tenfold cross validation than LR, NB, RF, SVM, KNN models [9]. Stephen et al. [10], analyzed the comparison of kyphosis prediction and attained benchmark results [10]. Rajasekaran et al. [11], proposed classification for morphology of column deficiency, curve magnitude, and flexibility and presents solution for selection of matching osteotomy [11].

Chauhan et al. [12], analyzed deep neural network for kyphosis classification and utilized fivefold and tenfold cross validation and achieved an high accuracy for 87.87% for fivefold cross validation and 87.67% for tenfold cross validation [12]. Kim et al. [13], diagnosed the proximal junctional kyphosis and categorized the risk factors based on patient related, radiological and surgical related and suggest preventive method for PJK [13, 14].

## 4 Workflow

The workflow of kyphosis classification is depicted in the following figure, utilizing a dataset sourced from Kaggle for experimentation. Although kyphosis is typically not severe, it can significantly impact one's normal body structure. Severe kyphosis can result in discomfort and lead to physical disfigurement. This paper aims to investigate the status of kyphosis in human subjects (Fig. 1).



**Fig. 1** Overall workflow

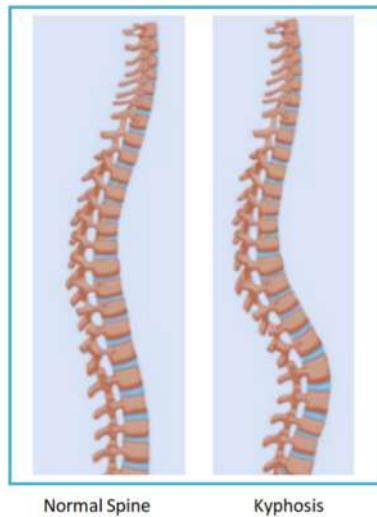
## 5 Dataset

It was taken from kaggle, and it contains 87 rows and 4 columns, it's a simple and small dataset, hence dataset improved by feature engineering and transformation. Feature scaling carried out by one hot encoding. The “Age” parameter refers to the age of patients measured in months, “Number” indicates the quantity of vertebrae affected, and “Start” denotes the initial vertebra (the uppermost) targeted for surgery. Following figures depict the normal and kyphosis spine.

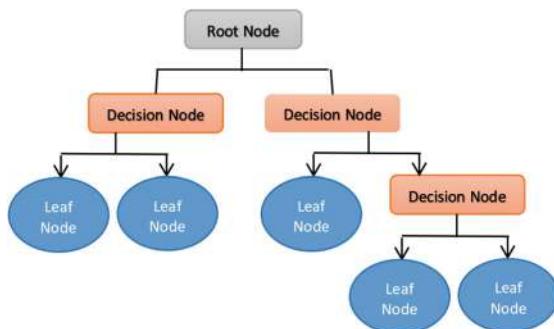
## 6 Exploratory Data Analysis (EDA)

It is a fundamental step in data preprocessing, crucial for identifying significant patterns and correlations among features. This process aids in selecting the most suitable model by uncovering valuable insights within the dataset. One hot encoding eliminates ordinal relationships in EDA by converting categorical variables into binary vectors. By making the data format understandable for algorithms, this conversion enhances analysis by facilitating better decision-making (Fig. 2).

**Fig. 2** Normal and kyphosis spine



**Fig. 3** Basic structure of decision tree



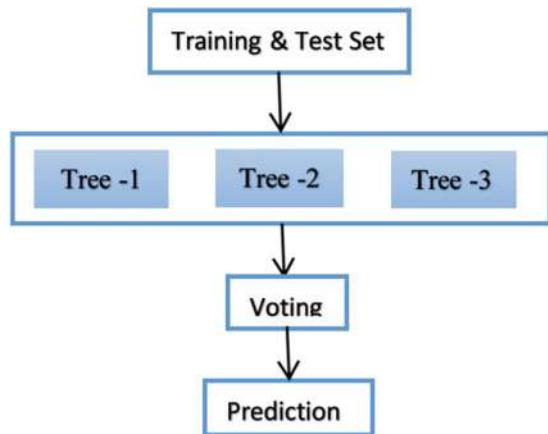
## 7 Decision Tree Classifier

Using input data as a guide, a decision tree classifier builds a tree structure, dividing important features at each node. It handles both numerical and categorical values efficiently as it moves from the root to the leaf nodes. Following figure shows the basic model of decision tree (Fig. 3).

## 8 Random Forest Classifier

The random forest classifier, operating similarly to decision trees, enhances model generalization and classification accuracy by leveraging an ensemble approach. Random Forest classifiers are able to handle complicated data with ease and prevent

**Fig. 4** Basic model of Random Forest



over fitting. They can handle unbalanced datasets well because they automatically choose pertinent features and function independently for each decision tree (Fig. 4).

## 9 Convolution Neural Network (CNN)

It is one of the basic and strong deep learning model and it build by multiple layers. The layers convolution, pooling, padding, Activation function and fully convolution layer. Due to its unsupervised learning nature, it automatically learned hierarchical features from input. In Convolutional Neural Networks (CNNs), preprocessing involves enhancing the dataset through synthetic data creation.

**Input Layer:** Receives input data, typically images.

**Convolutional Layer:** Utilizes filters to capture strong patterns, producing feature maps through element-wise multiplication with the input image followed by summation.

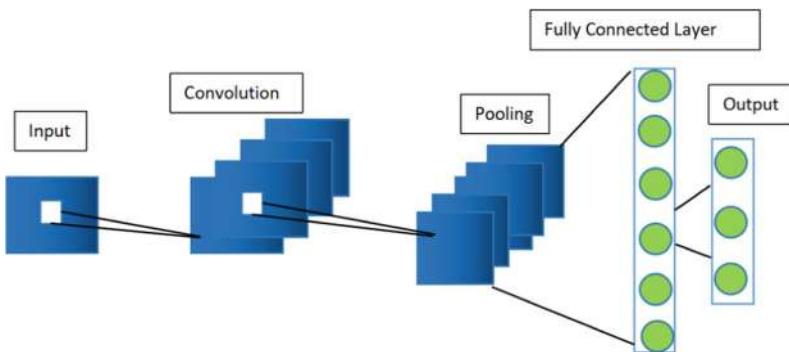
**Activation Function:** Enhances learning by introducing non-linearity to detect complex patterns.

**Pooling Layer:** Reduces spatial information while preserving important features.

**Fully Connected Layer:** Integrates high-level features extracted by convolutional layers.

**Output Layer:** Produces final predictions or classifications based on learned features.

Following figure shows the basic architecture of CNN (Fig. 5).



**Fig. 5** Base architecture of CNN

## 10 Evaluation Metrics

Several assessment metrics were used to evaluate the model's performance.

**Confusion Matrix:** A classification model's performance is evaluated using the Confusion Matrix, which compares expected and actual outcomes. It provides information on recall, accuracy, precision, and F1-score and consists of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

**Accuracy:** The number of correctly predicted instances divided by the total number of instances in the dataset is used to calculate accuracy, a metric used to assess a classification model's performance.

## 11 Result and Discussion

Three AI-driven algorithms were used in this study to classify cases of kyphosis. The study started with Exploratory Data Analysis (EDA) and moved on to employing Random Forest, Decision Tree, and Convolutional Neural Network (CNN) models for implementation and evaluation. The findings are presented.

The features included in the dataset are shown in Fig. 6.

The dataset has the following representation following one hot encoding (Fig. 7).

The heat map in the figure illustrates correlation; a diagonal value of 1 denotes perfect correlation between a variable and itself (Fig. 8).

The distribution of people with kyphosis in various age groups is shown in the following figure (Fig. 9).

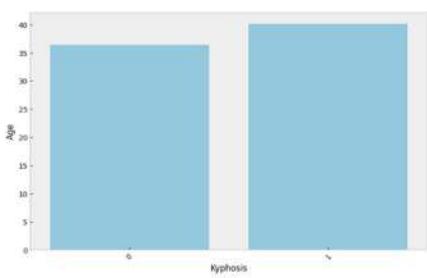
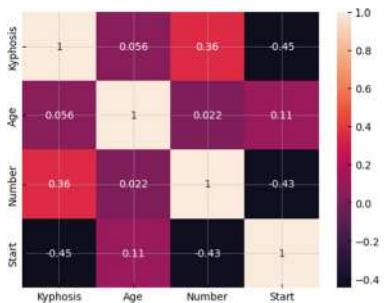
In the analysis of figure (a), it was observed that when kyphosis is plotted on the x-axis and age on the y-axis, the rate of kyphosis presence surpasses its absence. In figure (b), where age is plotted on the x-axis and kyphosis on the y-axis, it is noted

**Fig. 6** Feature in the dataset

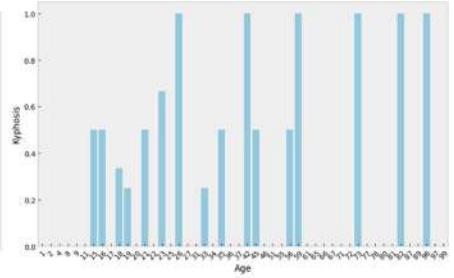
	Kyphosis	Age	Number	Start
0	absent	71	3	5
1	absent	67	3	14
2	present	18	4	5
3	absent	2	5	1
4	absent	1	4	15

**Fig. 7** After one hot encoding

	Kyphosis	Age	Number	Start
0	0	71	3	5
1	0	67	3	14
2	1	18	4	5
3	0	2	5	1
4	0	1	4	15

**Fig. 8** Correlation heat map

(a)



(b)

**Fig. 9** a and b kyphosis in various age group

that individuals aged 26, 35, 42, 59, 72, 83, and 96 years are primarily affected by kyphosis.

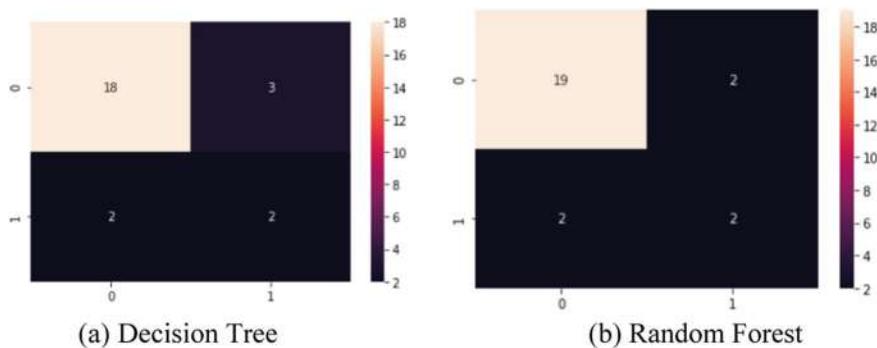
The following figures shows the confusion matrix of decision tree classifier and random forest classifier and its accuracy (Figs. 10 and 11).

Based on the aforementioned figure, this study finds that Random Forest outperforms then Decision Tree in terms of accuracy.

Following table and figure show the comparison of existing and proposed model (Table 1, Fig. 12).

Following figure shows the layer structure of CNN (Fig. 13).

Following figure show the Accuracy and loss performance of CNN (Fig. 14).



**Fig. 10 a and b** Confusion matrix of decision tree and Random Forest

**Fig. 11 a and b** Accuracy of decision tree and Random Forest

	precision	recall	f1-score	support
0	0.90	0.86	0.88	21
1	0.40	0.50	0.44	4
accuracy			0.80	25
macro avg	0.65	0.68	0.66	25
weighted avg	0.82	0.80	0.81	25

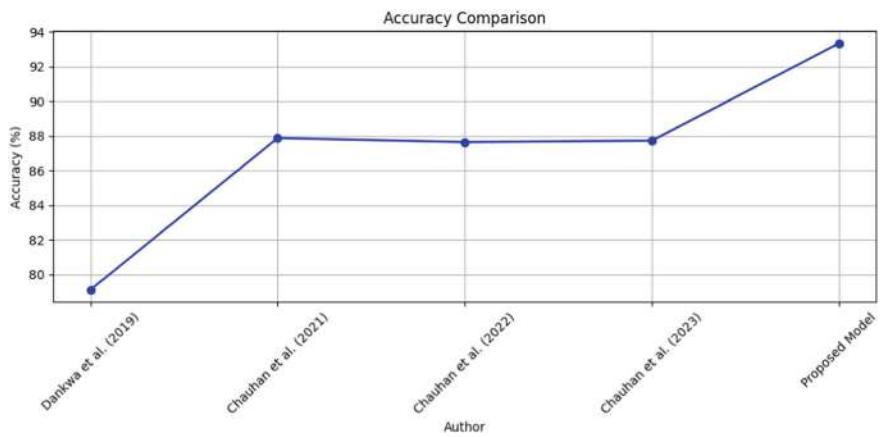
a) Decision Tree

	precision	recall	f1-score	support
0	0.90	0.90	0.90	21
1	0.50	0.50	0.50	4
accuracy			0.84	25
macro avg	0.70	0.70	0.70	25
weighted avg	0.84	0.84	0.84	25

(b) Random Forest

**Table 1** Comparison of existing and proposed model

Author	Methods	High accuracy (%)
Dankwa et al. [4]	FR, SVM, ANN	ANN-79.12
Chauhan et al. [12]	DNN	87.87
Chauhan et al. [3]	RF, SVM, KNN, DNN	DNN-87.64
Chauhan et al. [9]	DNN, LR, NB, RF, SVM, KNN	DNN-87.72
Proposed model	CNN	93.33

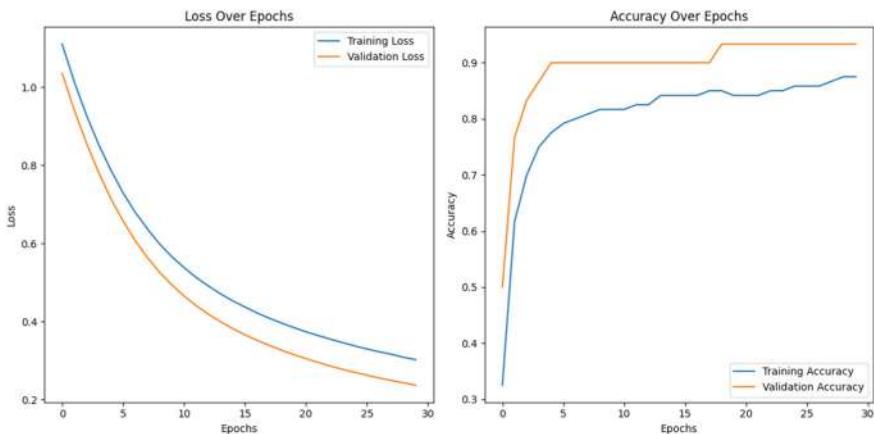
**Fig. 12** Comparison of existing and proposed model

From the above figure, it is observed that the CNN achieved an accuracy of 93.33% and loss of 0.75% only.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 62, 62, 32)	896
max_pooling2d (MaxPooling2D)	(None, 31, 31, 32)	0
conv2d_1 (Conv2D)	(None, 29, 29, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(None, 14, 14, 64)	0
conv2d_2 (Conv2D)	(None, 12, 12, 128)	73856
max_pooling2d_2 (MaxPooling2D)	(None, 6, 6, 128)	0
flatten (Flatten)	(None, 4608)	0
dense (Dense)	(None, 128)	589952
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 10)	1290
<hr/>		
Total params: 684490 (2.61 MB)		
Trainable params: 684490 (2.61 MB)		
Non-trainable params: 0 (0.00 Byte)		

---

**Fig. 13** Layer structure of CNN



**Fig. 14** Accuracy and loss performance of CNN

## 12 Conclusion

Kyphosis analysis is non-caring disease in many countries. Even people has no awareness about it and the challenges faced by the physicians. Hence, this research experiment a kyphosis analysis to help the medical diagnostics. This study efficiently utilized the Exploratory Data Analysis, and one hot encoding for preprocessing and implement the various machine learning and deep learning algorithms for classification of kyphosis analysis. The machine learning algorithms Random Forest and Decision Tree utilized and Decision Tree achieved an accuracy of 80%, Random Forest attained an accuracy of 84% and deep learning method CNN achieved an accuracy of 93.33%, and it is considerably higher than ML algorithms. This study suggests that CNN is efficient model for kyphosis classification.

## 13 Future Enhancement

The kyphosis analysis need more researcher's involvement to create a standard dataset.

Investigate this model with other spinal conditions and musculoskeletal disorders.

## References

1. Hussein YY, Khan MM (2023) Using artificial intelligence to predict the development of kyphosis disease: a systematic review. *Cureus* 15(11):e48341. <https://doi.org/10.7759/cureus.48341>. PMID: 38060748; PMCID: PMC10698623
2. Chatter P, Ramana D, Suzain S, Latha P (2021) Prediction of kyphosis disease using machine learning. [https://doi.org/10.1007/978-981-15-9712-1\\_30](https://doi.org/10.1007/978-981-15-9712-1_30)
3. Chauhan A, Gupta A, Garg R (2022) Comparative analysis of supervised machine and deep learning algorithms for kyphosis disease detection. <https://doi.org/10.21203/rs.3.rs-2091003/v1>
4. Dankwa S, Zheng W (2019) Special issue on using machine learning algorithms in the prediction of kyphosis disease: a comparative study. *Appl Sci* 9(16):3322
5. Passias PG, Vasquez-Montes D, Poorman GW, Protopsaltis T, Horn SR, Bortz CA et al (2018) Predictive model for distal junctional kyphosis after cervical deformity surgery. *Spine J* 18(12):2187–2194
6. Lafage R, Fourman MS, Smith JS, Bess S, Shaffrey CI, Kim HJ et al (2023) Can unsupervised cluster analysis identify patterns of complex adult spinal deformity with distinct perioperative outcomes? *J Neurosurg Spine* 38(5):547–557
7. Cina A, Galbusera F (2024) Advancing spine care through AI and machine learning: overview and applications. *EFORT Open Rev* 9(5):422–433. <https://doi.org/10.1530/EOR-24-0019>. PMID: 38726988; PMCID: PMC11099586
8. Ren G, Yu K, Xie Z, Wang P, Zhang W, Huang Y et al (2022) Current applications of machine learning in spine: from clinical view. *Global Spine J* 12(8):1827–1840
9. Chauhan AS, Gupta AK, Garg RR (2023) Comparative analysis of supervised machine and deep learning algorithms for kyphosis disease detection, published by 17 April 2023. <https://doi.org/10.3390/app13085012>
10. Dankwa S, Zheng W (2019) The prediction of kyphosis disease: a comparative study, special issue on using machine learning algorithms. *Appl Sci* 2019. <https://doi.org/10.3390/app9163322>
11. Rajasekaran S, Rajoli SR, Aiyer SN, Kanna R, Shetty AP (2018) A classification for kyphosis based on column deficiency, curve magnitude, and osteotomy requirement. *J Bone Joint Surg* 100(13):1147–1156. <https://doi.org/10.2106/JBJS.17.01127>
12. Chauhan AS (2021) Kyphosis disease prediction using deep neural networks. *Int J Eng Res Technol (IJERT)* 2278–0181. <https://doi.org/10.17577/IJERTCONV9IS05026>
13. Kim HJ, Yang JH, Chang DG, Suk SI, Suh SW, Kim SI, Song KS, Park JB, Cho W (2021) Proximal junctional kyphosis in adult spinal deformity: definition, classification, risk factors, and prevention strategies. *Asian Spine J* 2022 16(3):440–450. <https://doi.org/10.31616/asj.2020.0574>. Epub Apr 30. PMID: 33910320; PMCID: PMC9260397
14. Kim HJ, Iyer S (2016) Proximal junctional kyphosis. *J Am Acad Orthop Surg* 24(5):318–26. <https://doi.org/10.5435/JAAOS-D-14-00393>. PMID: 26982965

# An Enhanced Lightweight Authentication Scheme Based on Three Factors for WSN



Shilpi Sharma and Bijendra Kumar

**Abstract** Wireless sensor networks (WSN) are becoming increasingly prevalent because of their diverse applications, which include smart homes, automated production, healthcare, and environmental monitoring. In general, a WSN consists of a gateway node (GN), a user, and a sensor node (SN), that has restricted resources in smart devices. SNs are utilized in a wide range of industries and collect huge volumes of real-time data. GN manages the data acquired by installed SNs in order to deliver legitimate services for consumers. Using a portable device, the user may effortlessly control these smart products from anywhere. Because of their limited storage capacity, embedded systems make it difficult to establish comprehensive authentication methods. When implementing intelligent physical devices for smart applications, system management must include strong authentication, a lightweight security mechanisms, and appropriate consideration for user convenience. Thus, for secure authentication, an Enhanced Lightweight Authentication (ELA) Scheme based on three factors is presented in this approach. In general, authentication process in WSN has three phases that are pre-deployment, registration, and authentication. In Yu and Park, 2020 had presented password change process to enhance the security. However, they need to improve security against smartcard loss and session key leak. Thus, in this paper, we describe the session key update, user revocation, and re-registration phases. The proposed scheme's performance is evaluated with regard of communication cost, energy cost, and delivery ratio.

**Keywords** Wireless sensor networks · Authentication · Security · Lightweight · Three factors authentication

---

S. Sharma (✉) · B. Kumar

Department of Computer Science and Engineering, NSUT, Delhi, India

e-mail: [shilpi.sharma.cs19@nsut.ac.in](mailto:shilpi.sharma.cs19@nsut.ac.in)

B. Kumar

e-mail: [bizender@nsut.ac.in](mailto:bizender@nsut.ac.in)

## 1 Introduction

WSNs have received plenty of attention in recent years due to advancements in wireless communication along with sensor minimization technology [1, 2], as well as their practical application in a different environments such as the industrial Internet of Things (IIoT) [3], medical services [4], and smart homes [5]. A device linked to the IoT gathered data from its surroundings and transferred it to a server. Among the sensors, gateways, and users that make up a WSN, the communication consists of SNs and gateways. There are a variety of industrial environments where it is becoming increasingly necessary to monitor and collect data from. Basically, IIoT is a subset of IoT, and it concentrates on the important needs of applications in industry, like health surveillance, the agricultural sector, military, manufacturing, and consumer applications [6]. There is a significant role for WSNs to play in the IIoT when it comes to creating smart environments within the systems. Wireless SNs collect the data all the time, and then transmit this data to a database in order to be processed by the users, who then access the corresponding database in order to retrieve the desired data.

In other cases, users of a WSN may require data in real time at a certain period, for instance, during military surveillance as well as vehicle surveillance or tracking [7]. When these circumstances exist, users must communicate directly to wireless SNs so as to collect information. Critical data must also be sufficiently protected against unauthorized access along with illegal recording because WSNs are frequently used in threatening and unsupervised environments for a variety of purposes, including target tracking and battlefield surveillance. This makes them extremely concerning, especially in these situations [8]. Although safe information sharing across IoT participants is challenging due to the unsecured environment of wireless channels and the limited in resources features of SNs, several more challenges are to be expected. Thus in order to authenticate the participants of the key agreement, mutual authentication becomes a crucial security mechanism [9, 10].

In order to ensure that the above requirements in terms of security are met, a mechanism should be employed that is efficient. In the year 2020, Yu and Park [11] presented a password change process to enhance the security of passwords. It has been stated that the technique devised by Yu and Park protects users from numerous assaults while also providing them with anonymity, untraceability, and active authentication. Despite this, businesses still need to tighten their security to prevent smart-card losses and session key leaks. Furthermore, their technique is unsuited for WSN situations because to the substantial communication along with computation costs involved. To address the problem, we offer an Enhanced Lightweight Authentication Scheme Using Three Factors for Wireless Sensor Networks that considers the efficiency of smart devices while also boosting the security level of Yu and Park's method.

The primary contributions of our paper are outlined below:

- We present a secure and lightweight authentication method for WSN that addresses security issues with Yu and Park's approach by using secret parameters and biometrics.
- In this work, we present session key update, user revocation and re-registration phase to improve security against smartcard loss and session key leak.
- We analyze an informal security analysis to confirm that ELA provides secure mutual authentication.
- The suggested ELA delivers greater security and more features while having lower computational, communication, and storage overheads than previous systems.

## 2 Literature Review

To overcome the challenges in Sharif et al.'s scheme, Sahoo et al. [12] introduced an improved three-factor-based data transmission authentication system (TDTAS). They looked at the system used by Sharif et al. and found a number of security holes, such as a breach of user anonymity, ineffective login and password update procedures, and computation of session keys. They introduced a fuzzy extractor and a TDTAS, an enhanced ECC-based authentication technique. Additionally, a phase for the insertion of SNs and SC revocation was included in this system. Formal security analysis was performed using the Real-or-Random (RoR) paradigm, which guaranteed the security of the scheme's session key. A limitation of this scheme is associated with computational cost, which needs to be reduced further.

For WMSN configurations, Ali et al. [13] created an improved three-factor remote user authentication system. They looked at Amin et al.'s method and found a number of security flaws, such as an ID guessing attempt, an established session key short-term data threat, a user ID attack, and an offline password guessing threat. The authors created this method to address the previously identified security concerns. We verified the security of the protocol by applying the Burrows-Abadi-Needham (BAN) logic. A paradigm for safely generating session keys and mutual authentication is the BAN logic. Furthermore, comparing the performance of this scheme to other similar existing systems revealed greater complications in terms of communication and computing expenses.

Yu et al. [14] developed a safe and lightweight three-factor authentication technique designed for IoT-enabled smart home scenarios to solve flaws in Kaur and Kumar's protocol. Their suggested AKA method significantly reduces security concerns including impersonation as well as session key leak attacks while maintaining mutual authentication, anonymity, and privacy. The protocol underwent thorough formal security examinations, including simulation-based review with AVISPA for complete security analysis against multiple threats, as well as mathematical scrutiny with the ROR model to ensure session key management integrity. These validations underscore the protocol's robustness and suitability for ensuring secure communication and user privacy in IoT-based smart home environments.

Luo et al. [15] established a unique lightweight three-factor authentication strategy for WSN that can survive a variety of threats while providing superior operating efficiency over recently introduced schemes. They first examined Gope et al.'s approach, identifying various problems with it and other two-factor authentication techniques, including vulnerabilities to privileged insider attacks and clone card attacks. They suggested a ground-breaking three-factor authentication method that combines smart cards for authorized users, passwords, and biometrics to mitigate these worries. This innovative approach fits the requirements for mutual authentication and key agreement protocols, allowing users to securely access real-time data in WSNs. Both formal and informal security evaluations revealed that their method had superior security features and a higher security level.

Xue et al. [16] developed a lightweight multifactor authentication protocol for multi-gateway WSN utilizing hash functions and XOR operations. They first reanalyzed Guo et al.'s protocol, identifying various flaws and disadvantages, including a lack of repairability, poor consideration of biometric factors, sensitivity to offline password guessing attacks, and the absence of forward secrecy. They suggested a novel lightweight three-factor authentication as well as key agreement technique depends on the 12-Criteria framework that is designed for multi-gateway scenarios. Their approach included biometric traits that were confirmed with a fuzzy extractor, as well as the honey\_list mechanism for effective smart card logout. Extensive formal as well as informal security studies confirmed that the suggested approach is proper and secure. Comparative analyses with related studies revealed that this novel strategy achieved the optimal balance among security and effectiveness.

Yu et al. [11] created the Secure and Lightweight Three-Factor-Based User Authentication Protocol for Wireless Sensor Networks (SLUA-WSN). They found serious security problems in Mo and Chen's method, like session key disclosure and vulnerability to masquerade attacks, as well as a failure to ensure anonymity, untraceability, and authentication. To address these concerns, they created the SLUA-WSN, which uses biometric as well as secret parameters to improve security. The protocol successfully mitigates a variety of vulnerabilities, which includes SN acquisition, masquerade, and privileged insider assaults. They proved that SLUA-WSN offers safe mutual authentication for users, GNs, including SNs utilizing BAN logic. The security of the protocol was confirmed by formal security evaluations like the AVISPA simulation and the ROR model. Comparative performance assessments demonstrated that SLUA-WSN provides improvements in communication, computation, and storage overheads compared to existing schemes.

Kwon et al. [17] developed a Secure Three-Factor-Based Mutual Authentication Scheme with a Physical Unclonable Function (PUF) for Wireless Medical Sensor Networks. This technique was created to identify the security issues discovered in Masud et al.'s approach. Their suggested three-factor method, which incorporates biometrics and PUF, protects against numerous assaults while also offering perfect forward secrecy, anonymity and mutual authentication. To validate their scheme's security properties, they performed comprehensive evaluations using informal approaches, BAN logic, the RoR model, and AVISPA simulations. These

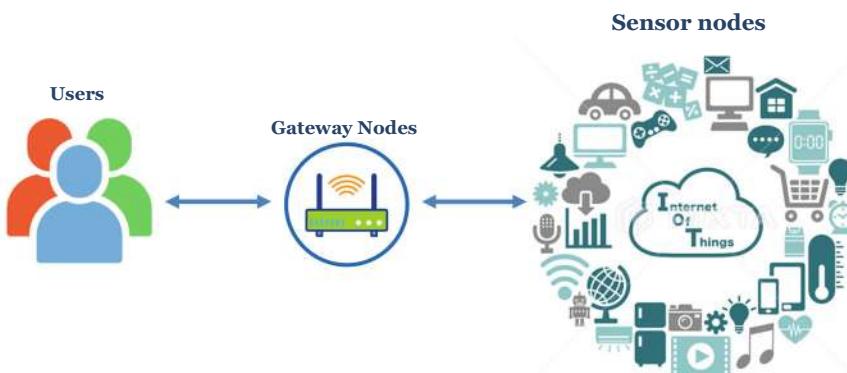
evaluations confirmed the robustness and effectiveness of their scheme in maintaining secure communication within wireless medical sensor networks.

Khemissa and Tandjaoui [18] et al. developed a new lightweight authentication mechanism designed for resource-constrained environments. Through mutual authentication, this method secured communication between the sensor as well as the remote user. This method uses Keyed-Hash message authentication, exclusive-or operations, and nonces to guarantee the integrity of the many exchanges. The approach being suggested is unique in that it provides authentication with lower energy consumption and concludes with a session key agreement between the remote user and the SN. They examined this strategy through a performance and security analysis. The outcomes demonstrated that this strategy resists different kinds of attacks and conserves energy.

### 3 Proposed Methodology

#### 3.1 System Model

In 2020, Yu and Park [11] had presented password change process to enhance the security in Three-Factor-Based User Authentication Protocol. However, they need to improve security against smartcard loss and session key leak. This work presents ELA approach based on three-factor that improves security by incorporating additional phases. Figure 1 illustrates the system model, which includes three entities: The user, the SN, and the GN. At first, the user communicates GN to establish the key agreement among itself and the SN. The SN uses a GN for mutual authentication that ensures the user is authentic. This approach supports mutual authentication and key agreement among users and SNs.



**Fig. 1** System model

### 3.2 Enhanced Lightweight Authentication Scheme

This paper addresses the security problems found in [11] by proposing an enhanced lightweight and secure user authentication protocol for WSN. As shown in Fig. 1, this protocol is composed of three entities such as the user ( $u_i$ ), SN<sub>j</sub>, and GN. Pre-deployment, user registration, authentication, password updating, session key updating, user revocation, and re-registration are the six phases of the suggested approach. The parts that follow go into great depth on each phase of the suggested method (Table 1).

#### 3.2.1 Pre-deployment Phase

This phase outlines the process by which a SN<sub>j</sub> registers with the GN.

- Step 1: GN chooses a distinct ID SId<sub>j</sub> for sensors and calculates  $Z_j = h(SId_j || K_{GN})$ . Finally, GN transmits {SId<sub>j</sub>, Z<sub>j</sub>} to the SN<sub>j</sub> via secured communication.
- Step 2: After getting the messages, the SN<sub>j</sub> stores them in safe memory.

#### 3.2.2 User Registration Phase

During the phase of user registration, a user  $u_i$  should register at the GN to gain access to various services. The process ensures that the user's ID, biometric data, and password are securely incorporated into the authentication system.

**Table 1** Notations

Notations	Descriptions
$u_i$	$i$ th user
SN <sub>j</sub>	$j$ th SN
GN	GN
SId <sub>j</sub>	SN <sub>j</sub> 's ID
Z <sub>j</sub>	Secret key of SN <sub>j</sub>
$h(\cdot)$	Hash function
$K_{GN}$	Master key of GN
Id <sub>i</sub>	ID of $u_i$
bio <sub>i</sub>	Biometric of $u_i$
Pwd <sub>i</sub>	Password of $u_i$
T <sub>s</sub>	Timestamps
Sk	Session key
	Concatenation
$\oplus$	XOR

- Step 1:  $u_i$  inserts the  $\text{Id}_i$ ,  $\text{Pwd}_i$  and imprints biometric  $\text{bio}_i$ . After that, the  $u_i$  calculates  $\text{Gen}(\text{bio}) = \langle q_i, p_i \rangle$  and  $\text{MPwd}_i = h(\text{Pwd}_i || q_i)$  and transmits  $\{\text{Id}_i, \text{MPwd}_i\}$  to the GN in a secured communication.
- Step 2: When a message is received, the GN produces a random nonce  $q_n$  and computes  $\text{MId}_i = h(\text{Id}_i || h(K_{GN} || q_n))$ ,  $Z_i = h(\text{MId}_i || q_n || K_{GN})$ ,  $R_i = h(\text{MId}_i || \text{MPwd}_i) \oplus Z_i$  and  $V_i = h(\text{MPwd}_i || Z_i)$  and saving  $\{q_n\}$  in a secured database. A smart card is then issued to the  $u_i$  with  $\{R_i, V_i, \text{MId}_i\}$  stored in it.

### 3.2.3 Authentication Phase

This phase is crucial for establishing a secure session key among the registered user  $u_i$ , the GN and  $\text{SN}_j$ .

- Step 1:  $u_i$  first puts the smart card (Sc) and then enters  $\text{Id}_i$  and  $\text{Pwd}_i$ . After that, the  $u_i$  imprints  $\text{bio}_i$  and also calculates  $q_i = \text{rep}(\text{bio}_i, p_i)$ ,  $\text{MPwd}_i = h(\text{Pwd}_i || q_i)$ ,  $Z_i = h(\text{MId}_i || \text{MPwd}_i) \oplus R_i$ , and  $V_i^* = h(\text{MPwd}_i || Z_i)$  and then verifies  $V_i^* = V_i$ . When the condition is true, the  $u_i$  produces a random nonce  $q_u$  along with a timestamp  $Ts_1$ . Then  $u_i$  computes  $M_1 = Z_i \oplus q_u$ ,  $C\text{Id}_i = (\text{Id}_i || S\text{Id}_j) \oplus h(\text{MId}_i || q_u || Z_i)$ , and  $M_{ug} = h(\text{Id}_i || q_u || Z_i || Ts_1)$ , and sends  $\{M_1, \text{MId}_i, C\text{Id}_i, M_{ug}, Ts_1\}$  to the GN in an insecure channel.
- Step 2: When a message is received, the GN verifies the authenticity of  $Ts_1$ , after computes  $Z_i = h(\text{MId}_i || q_n || K_{GN})$ ,  $q_u = M_1 \oplus Z_i$ ,  $(\text{Id}_i || S\text{Id}_j) = C\text{Id}_i \oplus h(\text{MId}_i || q_u || Z_i)$  and  $M_{ug}^* = h(\text{Id}_i || q_u || Z_i || Ts_1)$  and then, verify  $M_{ug}^* \stackrel{?}{=} M_{ug}$ . When the condition is valid, the GN calculates  $M_2 = (q_u || q_g) \oplus h(S\text{Id}_j || Z_j || Ts_2)$  and  $M_{gs} = h(\text{MId}_i || S\text{Id}_j || q_u || q_g || Z_j || Ts_2)$  and sends  $\{M_2, \text{MId}_i, M_{gs}, Ts_2\}$  to  $\text{SN}_j$ .
- Step 3: When a message is received, the  $\text{SN}_j$  verifies the authenticity of  $Ts_2$  and calculates  $(q_u || q_g) = M_2 \oplus h(S\text{Id}_j || Z_j || Ts_2)$  and then  $M_{gs}^* = h(\text{MId}_i || S\text{Id}_j || q_u || q_g || Z_j || Ts_2)$  and verifies  $M_{gs}^* \stackrel{?}{=} M_{gs}$ . When it is true, the  $\text{SN}_j$  produces a random nonce  $q_s$  along with timestamp  $Ts_3$  and computes  $M_3 = q_s \oplus h(q_u || S\text{Id}_j || Z_j || Ts_3)$ ,  $M_{sg} = h(q_s || q_g || S\text{Id}_j || Z_j || Ts_3)$ ,  $\text{Sk} = h(q_u || q_s)$  and  $M_{su} = h(\text{Sk} || q_s || q_g || \text{MId}_i || S\text{Id}_j)$  and then sends  $\{M_3, M_{sg}, M_{su}, Ts_3\}$  to the GN in a secure communication channel.
- Step 4: When a message is received, the GN verifies the authenticity of  $Ts_3$  and computes  $q_s = M_3 \oplus h(q_u || S\text{Id}_j || Z_j || Ts_3)$  and  $M_{sg}^* = h(q_s || q_g || S\text{Id}_j || Z_j || Ts_3)$ , verifies  $M_{sg}^* \stackrel{?}{=} M_{sg}$ . When it is true, the GN produces a timestamp  $Ts_4$  and then calculates  $\text{MId}_i^{\text{new}} = h(\text{Id}_i || h(K_{GN} || q_g))$ ,  $Z_i^{\text{new}} = h(\text{MId}_i^{\text{new}} || q_g || K_{GN})$ ,  $M_4 = (\text{MId}_i^{\text{new}} || Z_i^{\text{new}} || q_s || q_g) \oplus h(\text{MId}_i || Z_i || Ts_4)$ ,  $M_{gu} = h(q_u || q_g || \text{MId}_i || Z_i || Ts_4)$  and sends  $\{M_4, M_{su}, M_{gu}, Ts_4\}$  to  $u_i$ .
- Step 5: When a message is received, the  $u_i$  checks the validity of  $Ts_4$  and calculates  $(\text{MId}_i^{\text{new}} || Z_i^{\text{new}} || q_s || q_g) = M_4 \oplus h(\text{MId}_i || Z_i || Ts_4)$  and  $M_{gu}^* = h(q_u || q_g || \text{MId}_i || Z_i || Ts_4)$  and verifies  $M_{gu}^* \stackrel{?}{=} M_{gu}$ . If the condition is true,

the  $u_i$  calculates  $\text{Sk} = h(q_u||q_s)$  and  $M_{su}^* = h(\text{Sk}||q_s||q_g||\text{MId}_i||\text{SID}_j)$ , and verifies  $M_{su}^* \stackrel{?}{=} M_{su}$ . When the condition is true, then  $u_i$  calculates  $R_i^{\text{new}} = h(\text{MId}_i^{\text{MPwd}}||\text{MPwd}_i) \oplus Z_i^{\text{new}}$ , and  $V_i^{\text{new}} = h(\text{MPwd}_i||Z_i^{\text{new}})$  and replaces  $\{R_i, V_i, \text{MId}_i\}$  with  $\{R_i^{\text{new}}, V_i^{\text{new}}, \text{MId}_i^{\text{new}}\}$ . Then, the  $u_i$ , the GN and  $\text{SN}_j$  are successfully authenticated by each other.

### 3.2.4 Password Updating Phase

The Password Update Process allows an authorized user to securely update their password using their biometric data and a smart card.

- Step 1:  $u_i$  inputs  $\text{Id}'_i$  and  $\text{Pwd}'_i$  and prints biometric  $\text{bio}'_i$ . Then, the  $u_i$  calculates  $\text{Gen}(\text{bio}') = \langle q'_i, p'_i \rangle$  and  $\text{MPwd}'_i = h(\text{Pwd}'_i||q'_i)$  and then forwards  $\{\text{Id}'_i, \text{MPwd}'_i\}$  to the  $\text{Sc}$  via a secure communication channel.
- Step 2: When a message is received, the  $\text{Sc}$  calculates  $Z'_i = R'_i \oplus h(\text{MId}'_i||\text{MPwd}'_i)$  and  $V'_i = h(\text{MPwd}'_i||Z'_i)$  and transmits an authentication message to  $u_i$ .
- Step 3: When a message is received, the  $u_i$  selects a new  $\text{Pwd}_i^{\text{new}}$  and prints a new  $\text{bio}^{\text{new}}$ . After that,  $u_i$  computes  $\text{Gen}(\text{bio}^{\text{new}}) = \langle q_i^{\text{new}}, p_i^{\text{new}} \rangle$  and  $\text{MPwd}_i^{\text{new}} = h(\text{Pwd}_i^{\text{new}}||q_i^{\text{new}})$  and transmits  $\{\text{MPwd}_i^{\text{new}}\}$  to  $\text{SC}$  in a secure channel.
- Step 4: When a message is received, the  $\text{Sc}$  computes  $R_i^{\text{new}} = h(\text{MId}_i^{\text{new}}||\text{MPwd}_i^{\text{new}}) \oplus Z'_i$  and  $V_i^{\text{new}} = h(\text{MPwd}_i^{\text{new}}||Z'_i)$  and then replaces  $\{R'_i, V'_i\}$  with  $\{R_i^{\text{new}}, V_i^{\text{new}}\}$  successfully.

### 3.2.5 Session Key Updating Phase

A freshly generated random number for user  $q_u^{\text{new}}$  along with a timestamp  $Ts_5$  is needed to update the session key. GN verifies  $|Ts_5 - Ts_4| \leq Ts_d$  via user random number; if correct, GN forwards the updated message to the targeted  $\text{SN}_j$ . If  $|Ts_6 - Ts_5| \leq Ts_d$  is verified as true, GN generates a new secure session key  $\text{Sk}^{\text{new}} = h(\text{Id}_i||\text{SID}_j||(\text{Id}_i||q_u^{\text{new}})||h(q_{\text{sc}})||Ts_5||Ts_6)$  for additional communication.

### 3.2.6 User Revocation and Re-registration Phase

This phase is responsible for managing the secure removal of users from the network and their subsequent re-registration if necessary.

- Step 1:  $u_i$  transmits a new registration request to GN.  $u_i$  selects a new random number  $q_u^{\text{new}}$  and computes  $\text{MId}_i^{\text{new}} = h(\text{Id}_i||q_u^{\text{new}})$  and transmits it to the GN over a secure communication channel.
- Step 2: GN receives registration request from  $u_i$ , after that it calculates  $Z_i^{\text{new}} = h(\text{MId}_i||K_{\text{GN}})$ . GN stores  $Z_i^{\text{new}}$  into a new smartcard  $\text{Sc}_i^{\text{new}}$  and send to  $u_i$  over a secure communication channel.

- Step 3: After receiving  $Sc_i^{\text{new}}$ ,  $u_i$  chooses a new password  $\text{Pwd}_i^{\text{new}}$  and print a new biometric  $\text{bio}_i^{\text{new}}$  at sensor of specific terminal.  $Sc_i$  of  $u_i$  calculates  $\text{MPwd}_i = h(\text{Pwd}_i^{\text{new}} || q_u^{\text{new}})$ ,  $\text{Gen}(\text{bio}_i^{\text{new}}) = \langle q_i^{\text{new}}, p_i^{\text{new}} \rangle$ ,  $V_i^{\text{new}} = h(\text{MPwd}_i^{\text{new}} || q_i^{\text{new}})$ ,  $R_i^{\text{new}} = h(\text{Id}_i || \text{Pwd}_i^{\text{new}} || q_i^{\text{new}}) \oplus q_u^{\text{new}}$ ,  $Z_i' = Z_i^{\text{new}} \oplus h(\text{Id}_i || q_i^{\text{new}})$ .
- Step 4:  $Sc_i^{\text{new}}$  stores the newly computed parameter  $\{R_i^{\text{new}}, V_i^{\text{new}}, \text{MId}_i^{\text{new}}\}$  and issues it to the  $u_i$ .

## 4 Security Analysis

This part evaluated the security of the ELA scheme by informal security analysis.

### 4.1 Informal Security Analysis

An informal security assessment is used to evaluate the ELA system's security. We show that while offering safe authentication and anonymity, the ELA method is capable of withstanding following security threats.

#### 4.1.1 Masquerade Attack

The MA tries to imitate an authorized user via capturing messages delivered over an unprotected channel. Yet, the MA fails to generate request messages  $\{M_1, \text{MId}_i, \text{CId}_i, M_{ug}\}$  correctly under the proposed ELA scheme. The MA is unable to calculate the request messages since it lacks access to the user's true ID, biometrics, or random nonce. As a result, the ELA system can withstand disguise attacks.

#### 4.1.2 Replay Attack

When the MA conducts the replay attack using already transferred data in an insecure channel, the proposed ELA approach ensures the timestamp's freshness, even though the MA intercepted the request message  $\{M_1, \text{MId}_i, \text{CId}_i, M_{ug}, Ts_1\}$  from the prior session. Furthermore, the request packets are encrypted with a secret parameter  $Z_i$  and a randomly generated nonce  $q_u$ . Thus, the ELA system protects against replay assaults.

#### 4.1.3 SN Capture Attack

SNs usually appear in vacant or difficult areas, allowing the MA to rapidly capture them. But each  $SN_j$  possesses a unique  $SID_j$  along with secret parameter  $Z_j$ . Even though the MA captures certain SNs, impersonating the MA as another sensor is challenging. As a result, the MA has no authority to compromise additional agreements  $Sk$  made among the compromised  $u_i$  and the non-compromised parties  $SN_j$ . Thus, the ELA approach prevents SN capture assaults.

#### 4.1.4 Privileged Insider Attack

In this attack, an authorized user acquires access to the user's password stored in GN and uses it to impersonate the user on various systems. Yet, the user of the proposed ELA system only communicates  $Id_i$ ,  $MPwd_i$  with the GN when the registration is being processed. As a result, the ELA scheme guards against privileged attacks from inside by preventing the privileged insider from accessing the authentic user's actual password.

#### 4.1.5 Anonymity and Untraceability

It is assumed that the MA has the ability to intercept messages transferred between sessions and recover secret credentials from a smartcard. However, the MA is unable to locate a lawful user  $u_i$  because all transmitted messages have been revised each session, and  $\{R_i, V_i, MId_i\}$  messages within the proposed ELA scheme update with  $\{Z_i, R_i^{\text{new}}, V_i^{\text{new}}, MId_i^{\text{new}}\}$ . Furthermore, because  $Id_i$  of  $u_i$  is veiled by XOR and hash operations, the MA is unable to recover its true value. As a result, the ELA approach assures anonymity and untraceability by preventing the MA from extracting  $Id_i$  of  $u_i$  without a secret parameter  $Z_i$  along with a random nonce  $q_u$ .

#### 4.1.6 Mutual Authentication

In the ELA system, every entity successfully completes mutual authentication. After receiving authentication request messages,  $\{M_1, MId_i, CId_i, M_{ug}\}$  from the  $u_i$ , the GN verifies  $M_{ug}^* \stackrel{?}{=} M_{ug}$ . When the condition is true, the GN will authenticate it. After receiving the messages,  $\{M_2, MId_i, M_{gs}, Ts_2\}$  from the GN, the  $SN_j$  checks  $M_{gs}^* \stackrel{?}{=} M_{gs}$ . If it is correct, the  $SN_j$  authenticates the GN. Once getting the messages,  $\{M_3, M_{sg}, M_{su}, Ts_3\}$  from the  $SN_j$ , the GN verifies  $M_{sg}^* \stackrel{?}{=} M_{sg}$ . When the condition is true, the GN verifies the  $SN_j$ . After receiving the response messages  $\{M_4, M_{su}, M_{gu}, Ts_4\}$  from the GN, the  $u_i$  verifies the GN. Thus, the  $u_i$  and the  $SN_j$  and the GN are mutually authenticated as the MA unable produce exchanged messages exchanged messages  $\{M_{ug}, M_{gs}, M_{sg}, M_{su}\}$  successfully.

## 5 Results and Discussions

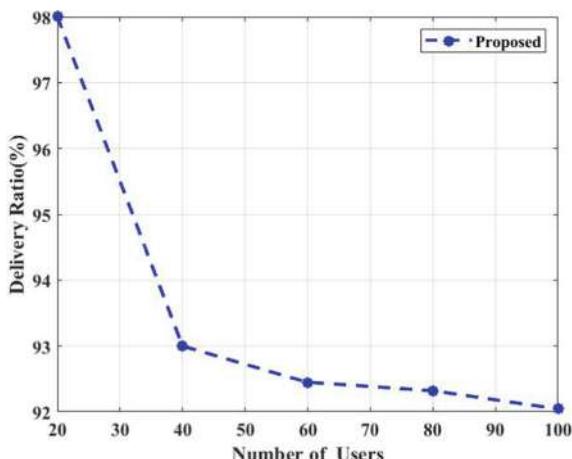
This section examines the efficacy of the proposed technique. The MATLAB tool is used to carry out the proposed method. The implementation requires a CPU-based computer system with 8 GB of RAM, a 256 GB of memory, and 2 GHz Intel Core i7 processor. We evaluate the performance of ELA scheme in terms of the delivery ratio, energy cost, and communication cost. These specifics are covered in the sections that follow.

Figure 2 depicts a comparison of the delivery ratio for the proposed the “ELA scheme” regularly achieves the highest delivery ratios, with values starting at 98% for 20 users, indicating almost all messages are successfully delivered and lowering to 93% for 40 users demonstrating a drop in efficiency with more users. As the pace increased, the “ELA scheme” achieved a lower delivery ratio, reaching 92% for 100 users.

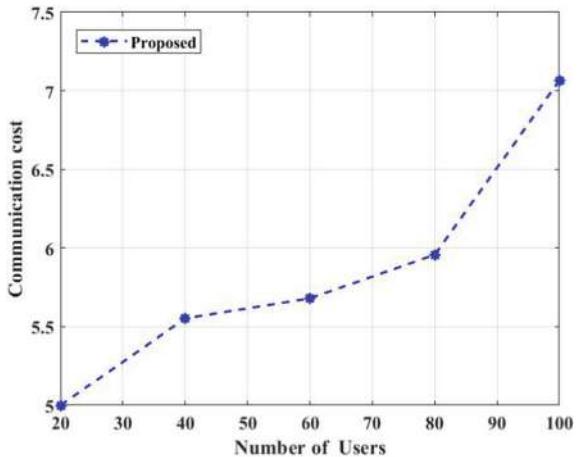
Figure 3 depicts a comparison of the Communication Cost for the proposed the “ELA scheme” regularly achieves the lowest Communication Cost. The proposed framework executes secure data transactions amid the quantity of data sent in the network at a low cost of transmission. For 100 users, our proposed ELA achieved 7.35 bits of communication cost.

Figure 4 depicts a comparison of the Energy Cost for the proposed the “ELA scheme” regularly achieves the lowest Energy Cost. The proposed framework executes secure data transactions amid the quantity of data sent in the network at a low cost of transmission. For 100 users, our proposed ELA achieved 24 units of energy cost.

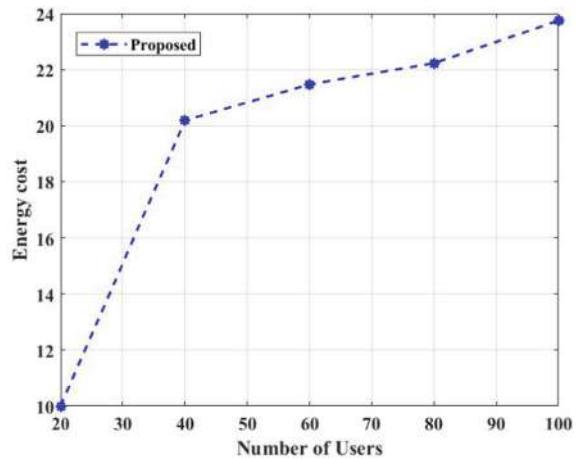
**Fig. 2** Performance analysis based on delivery ratio



**Fig. 3** Performance analysis based on communication cost



**Fig. 4** Performance analysis based on energy cost



## 6 Conclusion

In this study, we show that Yu and Park's approach has several security weaknesses, including smartcard loss and session key leak. Thus, for secure authentication, an Enhanced Lightweight Authentication (ELA) Scheme Based on Three Factors is presented in this approach. Thus, in this paper, we discussed session key updates, user revocation, and the re-registration step. The ELA Scheme prevents a wide range of threats, including SN capture, masquerade, including privileged insider attacks. We conducted security analysis to demonstrate that the proposed ELA Scheme supports secure mutual authentication across  $u_i$ , GN, and SN<sub>j</sub>. We also analyze ELA's delivery ratio, communication cost, and energy consumption. As a result, the

suggested ELA Scheme significantly improved security levels as compared to three-factor-based similar schemes while maintaining low computing and communication overheads by employing only XOR and hash operations. As a result, the proposed ELA Scheme outperformed previous schemes in terms of security and efficiency, making it appropriate for real-world WSN situations.

## References

1. Zhang R, Cui S, Zhao C (2020) Design of a data acquisition and transmission system for smart factory based on NB-IoT. In: Communications, signal processing, and systems: proceedings of the 2018 CSPS volume III: systems 7th. Springer Singapore, pp 875–880
2. Shi Y, Zhao Y, Xie R, Han G (2019) Designing a structural health monitoring system for the large-scale crane with narrow band IoT. In: 2019 IEEE 23rd international conference on computer supported cooperative work in design (CSCWD). IEEE, pp 239–242
3. Lara E, Aguilar L, Sanchez MA, Garcia JA (2020) Lightweight authentication protocol for M2M communications of resource-constrained devices in industrial Internet of Things. Sensors 20(2):501
4. Park K, Noh S, Lee H, Das AK, Kim M, Park Y, Wazid M (2020) LAKS-NVT: provably secure and lightweight authentication and key agreement scheme without verification table in medical internet of things. IEEE Access 8:119387–119404
5. Oh J, Yu S, Lee J, Son S, Kim M, Park Y (2021) A secure and lightweight authentication protocol for IoT-based smart homes. Sensors 21(4):1488
6. Malik PK, Sharma R, Singh R, Gehlot A, Satapathy SC, Alnumay WS, Pelusi D, Ghosh U, Nayak J (2021) Industrial Internet of Things and its applications in industry 4.0: state of the art. Comput Commun 166:125–139
7. Ali A, Ming Y, Chakraborty S, Iram S (2017) A comprehensive survey on real-time applications of WSN. Futur Internet 9(4):77
8. Pundir S, Wazid M, Singh DP, Das AK, Rodrigues JJ, Park Y (2019) Intrusion detection protocols in wireless sensor networks integrated to Internet of Things deployment: survey and future challenges. IEEE Access 8:3343–3363
9. Ezema E, Abdullah A, Mohd NFB (2018) Open issues and security challenges of data communication channels in distributed Internet of Things (IoT): a survey. Circ Comput Sci 3(1):22–32
10. Gopalakrishnan K (2020) Security vulnerabilities and issues of traditional wireless sensors networks in IoT. Princ Internet Things (IoT) Ecosyst Insight Parad 519–549
11. Yu S, Park Y (2020) SLUA-WSN: secure and lightweight three-factor-based user authentication protocol for wireless sensor networks. Sensors 20(15):4143
12. Sahoo SS, Mohanty S, Sahoo KS, Daneshmand M, Gandomi AH (2023) A three-factor-based authentication scheme of 5G wireless sensor networks for IoT system. IEEE Internet Things J 10(17):15087–15099
13. Ali R, Pal AK, Kumari S, Sangaiah AK, Li X, Wu F (2024) An enhanced three factor based authentication protocol using wireless medical sensor networks for healthcare monitoring. J Ambient Intell HumIzed Comput 1–22
14. Yu S, Jho N, Park Y (2021) Lightweight three-factor-based privacy-preserving authentication scheme for IoT-enabled smart homes. IEEE Access 9:126186–126197
15. Luo H, Wen G, Su J (2020) Lightweight three factor scheme for real-time data access in wireless sensor networks. Wireless Netw 26(2):955–970
16. Xue L, Huang Q, Zhang S, Huang H, Wang W (2021) A lightweight three-factor authentication and key agreement scheme for multigateway WSNs in IoT. Secur Commun Netw 2021(1):3300769

17. Kwon D, Park Y, Park Y (2021) Provably secure three-factor-based mutual authentication scheme with PUF for wireless medical sensor networks. Sensors 21(18):6039
18. Khemissa H, Tandjaoui D (2016) A novel lightweight authentication scheme for heterogeneous wireless sensor networks in the context of Internet of Things. In: 2016 Wireless telecommunications symposium (WTS). IEEE, pp 1–6

# EfficientDet with SAM on NC4K Dataset



Lavish Kumar and Shweta Meena

**Abstract** Camouflaged object detection (COD) has been a big challenge. It is a crucial task with numerous applications in various domains. It is contesting due to the resemblance between the object we are trying to detect and its background. Numerous methods and datasets have been introduced to tackle this problem and this domain has become one of the fastest growing domains of image processing. In this paper, we have implemented EfficientDet with SAM over a new dataset NC4K and have compared their results with some existing models. We have further analyzed the model's failures to identify limitations and proposed potential improvements for future work. This research contributes to the exploration of EfficientDet as well as SAM by comparing it with several models for COD and the effectiveness of the new NC4K dataset.

**Keywords** EfficientDet · Segment Anything Model (SAM) · Camouflaged Object Detection (COD) · NC4K

## 1 Introduction

### 1.1 Overview

Camouflaged object detection (COD) is a crucial and essential task in Visual Computing, specifically Computer Vision (CV) due to the great resemblance between the target entity and its surrounding environment. The traditional object detection

---

Shweta Meena is contributed equally to this work.

---

L. Kumar (✉) · S. Meena

Department of Software Engineering, Delhi Technological University (DTU), Delhi, New Delhi, India

e-mail: [kumar.lavish.0109@gmail.com](mailto:kumar.lavish.0109@gmail.com)

S. Meena

e-mail: [shwetameena@dtu.ac.in](mailto:shwetameena@dtu.ac.in)

techniques fail for this task due to the great resemblance between the targeted entity and its surrounding environment. Hence, image segmentation plays an important role in COS to overcome this challenge.

Image segmentation is a technique that is used to divide the image into separate segments, or pixel sets, and transform them into a representation that is easier to understand and study. Each segment is given a label based on its characteristics. Image segmentation is beneficial in the procedure of isolating the specific camouflaged object from the background.

## 1.2 Dataset

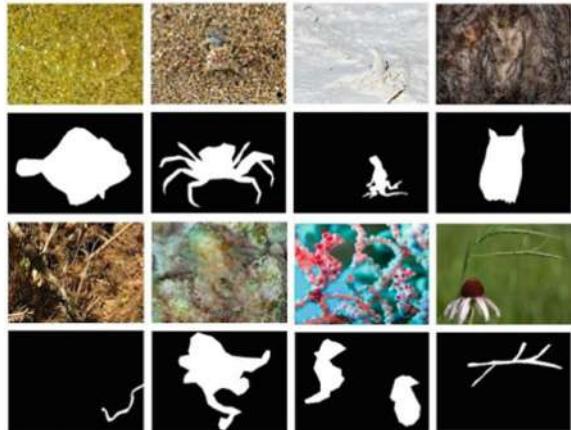
To do camouflaged object detection (COD), we require access to a large and diverse dataset for training and evaluation. Over the years, several datasets have been developed to address this problem, which results in the advancement of COD/COS research.

One of the most popular datasets for training COD models is the CAMO dataset [1] which was introduced in 2017. This dataset focuses specifically on the challenges related to Camouflaged Object Segmentation (COS), aiming to differentiate between the object and its background. It contains 1250 images that are split into training and testing sets, featuring both naturally camouflaged objects like animals and artificially camouflaged objects like man-made objects. However, the ground truth information provided was in COCO JSON format which gave limited information. Later, CAMO++ dataset [2] was introduced in 2021 which is likely larger than CAMO and provides hierarchical pixel-wise ground truths.

One of the biggest datasets for COD is COD10K [3] which is a well-established dataset introduced in 2020 and updated in 2022, frequently used for benchmarking COD algorithms. It contains 10,000 images, and the content of this dataset varies across different categories, including terrestrial animals, aquatic creatures, and even camouflaged objects created by humans. It offers rich ground truth information for each image, including Bounding Boxes, Categories and Attributes, Object and Instances Annotations, and Edge Annotations.

We have used the NC4K dataset [4] for this paper. This dataset is comparatively new and has nearly 4000 images. The images vary across various fields, from naturally camouflaged objects to artificially camouflaged objects. Its ground truth is similar to the CAMO++'s ground truth. As this is a new dataset, there is relatively less work done on this dataset. Some instances of the NC4K dataset can be seen in Fig. 1.

**Fig. 1** Image segmentation of camouflaged animals



### 1.3 Algorithm

For this paper, we have used EfficientDet [5]. EfficientDet is an object detection model known for its balance between accuracy and computational efficiency. This model was introduced by Google Research in 2020. Its strength lies in achieving high object detection accuracy while requiring less processing power compared to other state-of-the-art models. This makes it suitable for deployment on devices with limited resources. Some of its key features are compound scaling and BiFPN.

We have also used the Segment Anything Model (SAM) [6], which was introduced by Meta AI. It is a segmentation model designed for segmenting objects in images, even those unseen during the training phase. Its strength lies in zero-shot segmentation, which helps it to identify and segment objects without needing specific training data for those objects. Its components include Image Encoder, Prompt Encoder, and Mask Decoder.

## 2 Problem Statement

Camouflaged object detection (COD) plays an important role in various fields like marine biology, surveillance systems, and military operations. With the recent introduction of the NC4K dataset, we can try to achieve benchmarks for various COD algorithms. Due to its large size and diverse real-world camouflaged scenarios, it can be an efficient dataset. There is a lack of comprehensive research that can explore the full potential of the NC4K dataset.

This research gap slows the development of COD research. There exist many powerful algorithms that are capable of achieving high accuracy on the NC4K dataset, and they need to be tested. We have used EfficientDet and SAM on the NC4K dataset for the first time to address this gap. We can contribute to the future research of COD

by applying EfficientDet and SAM to NC4K and getting valuable insights. This analysis will provide a deeper understanding of how well the model handles various camouflage types and identify potential areas for improvement.

### 3 Preliminary and Background

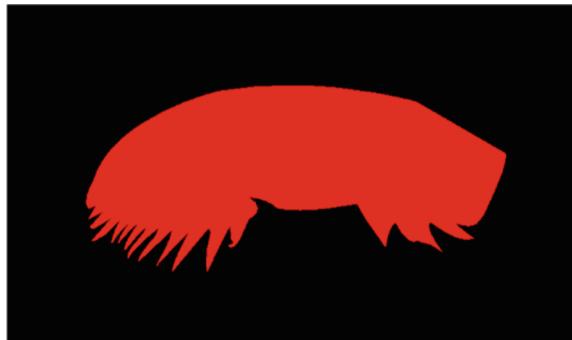
#### 3.1 NC4K

The Natural Camouflaged 4K dataset or NC4K dataset is a significant resource in the field of COD research. This is relatively a new dataset that comprises a diverse collection of 4121 high-resolution images gathered from the internet. These images include various natural settings, like forests, grassland, sea, mountainous regions, and deserts with a wide range of camouflaged objects like animals, marine creatures, insects, and man-made objects. One of its examples is visible in Figs. 2 and 3.

**Fig. 2** Picture of a fish from the NC4K dataset



**Fig. 3** Image of an instance of a fish from the NC4K dataset



For further research and development in COD, this dataset categorizes its images into three main groups, namely camouflaged objects, backgrounds, and non-camouflaged objects. Each image here is remarked with ground truth labels that indicate the presence of camouflaged objects. The ground truth includes segmented masks for each image.

One limitation of this dataset is that the images are collected from the internet so there can be a potential bias while training.

### 3.2 EfficientDet

EfficientDet is a model developed by Google Research for object detection. It is one of the most efficient models, as it prioritizes a balance between achieving high accuracy in object detection and maintaining computational efficiency. Hence, this model is suitable for deployment on platforms that offer limited resources.

EfficientDet is different from its predecessor EfficientNet [7]. EfficientNet addressed the challenge of scaling CNNs for better accuracy by introducing compound scaling which is a method to scale resolution, depth, and width simultaneously for optimal performance. The two most important key features of EfficientDet are compound scaling and BiFPN. Compound scaling is a technique that scales all aspects of the model architecture, including resolution, depth, width, etc., proportionally during the training phase. This makes sure that all the parts of the model contribute effectively to the final performance. BiFPN, or Bi-Directional Feature Pyramid Network, is a novel component of the model that works to improve feature extraction. It helps in maintaining the flow of information in both top-down and bottom-up directions within the network. With this, EfficientDet is capable of capturing a wider range of features that are important for accurate object detection. The conventional FPN aggregates multi-scale features in a top-down manner:

$$P_7^{\text{out}} = \text{Conv}(P_7^{\text{in}}) \quad (1)$$

$$P_6^{\text{out}} = \text{Conv}\left(P_6^{\text{in}} + \text{Resize}\left(P_7^{\text{out}}\right)\right) \quad (2)$$

$$P_3^{\text{out}} = \text{Conv}\left(P_3^{\text{in}} + \text{Resize}\left(P_4^{\text{out}}\right)\right) \quad (3)$$

The Equations (1), (2), and (3) were presented by Tan et al. in their paper [8].

There exist a variety of predefined EfficientDet models. These variants offer a trade-off between accuracy and efficiency. Smaller models like D0 and D1 prioritize efficiency for faster processing while larger models like D5 and D6 prioritize higher accuracy for more demanding tasks.

**Fig. 4** SAM being used for masking an image



### 3.3 SAM

Segment Anything Model or SAM, is a segmentation model that is built for identifying and segmenting objects within images. It was introduced by Meta AI in 2023. Traditional segmentation models require extensive training data for specific objects, but SAM comes with zero-shot segmentation capability. This is SAM's most significant strength. This helps SAM segment objects in an image without needing to train the model on those specific objects beforehand. This provides a great advantage for tasks involving unseen categories or rare objects. It can be seen in Fig. 4.

SAM provides prompt-based learning which means we can use natural language prompts to direct the process of segmentation according to ourselves. This feature allows greater control over the segmentation task.

There are three main components of SAM: Image Encoder, Prompt Encoder, and Mask Decoder. The Image Encoder extracts features from the input image. SAM uses a pre-trained Vision Transformer (ViT) for this purpose. ViTs are powerful models that are capable of capturing complex visual relationships within images. Prompt Encoder processes different types of prompts like text descriptions, bounding boxes, segmentation masks, etc., to understand the desired segmentation goals. Mask Decoder generates a segmentation mask for each object in the image based on the encoded image features and prompt information.

## 4 Related Works

Recently, there has been significant progress in the area of COD, which is progressed by the advancement in image segmentation, machine learning, and Computer Vision techniques. Multiple techniques and approaches have been suggested and implemented to address this challenge. Here, we have provided an overview of the existing literature on COD. We have focused on key techniques, datasets, and methodologies.

As well, we discuss works related to SAM and zero-shot learning. By examining the research in these areas, we have identified gaps and opportunities for further advancements in COD.

BASNet [9] and EGNet [10] are among the earliest models introduced for object detection in camouflaged backgrounds. They both are introduced in 2019. BASNet introduced a predictive refine architecture specifically designed for salient object detection that is boundary-aware. It uses a residual refinement module and a densely supervised Encoder-Decoder network. This helped BASNet to improve the boundary quality of salient object detection. BASNet uses a combined loss function with Intersection-over-Union (IoU) losses, Binary Cross Entropy (BCE), and Structural Similarity (SSIM) to guide the network in learning the conversion between input images and ground truth at multiple levels of hierarchy.

EGNet was introduced by Wang et al. to address the challenge of coarse object boundaries in salient object detection by upgrading the relationship between prominent entity information and salient edge information that is complementary. To continually describe these two types of complementing information in a single network, the model included an edge guidance network (EGNet).

SiNet [8] and PraNet [11] were introduced in 2020. They represented significant advancements in the fields of COD and medical image segmentation, respectively. SiNet addresses the challenge of COD by introducing a simple yet effective framework. Wang et al. developed a Search Identification Network (SINET) that provided a robust and general framework for COD that outperformed various benchmark object detection baselines on all tested datasets. This model leveraged the novel dataset COD10K which has 10,000 densely annotated images. The dataset covered camouflaged objects in diverse natural scenes across 78 object categories.

Fan et al. introduced the Parallel Reverse Attention Network or PraNet in 2020. PraNet was developed to accurately segment the polyp in colonoscopy images. This was an important model as it was used in the task of colorectal cancer detection and surgery. The PraNet used Parallel Partial Decoder (PPD) and Reverse Attention (RA) modules to improve segmentation accuracy. PraNet significantly improves segmentation accuracy with the help of continuous cooperation mechanisms between areas and boundaries. It is an efficient tool in the domain of medical image analysis as it is capable of doing real-time segmentation proficiently. It also has an advantage in terms of generalizability.

In 2021, SiNet-V2 [12] was introduced and it represented a significant advancement in camouflaged object detection. There are cases when objects perfectly blend in their background and that makes the task of object detection more demanding than standard object detection or segmentation to tackle this, SiNet-V2 was developed. The authors introduced the COD10K dataset which comprises 10,000 camouflaged object images across diverse real-world scenarios. There are a total of 78 categories ranging from animals to man-made objects. This dataset includes various annotations such as object categories, boundaries, challenging attributes, and instance-level annotations which makes it the largest and most comprehensively annotated COD dataset to date. SiNet-V2 is an upgraded version of SiNet and performs better on all the tested datasets. As a result, SiNet-V2 proposed future scope in the field of COD.

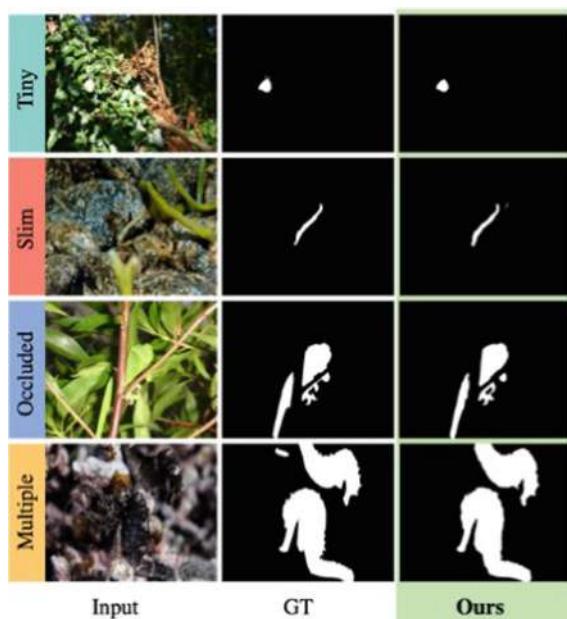
In 2021, Le et al. [13] introduced a novel task in the field of COD which is camouflaged instance segmentation that aims to decompose camouflaged regions in images into meaningful components. In order to support their task, the authors extended the CAMO dataset and introduced CAMO++ which is a dataset that substantially increases the quantity and diversity of images with pixel-wise ground truths in hierarchical order. By utilizing the CAMO++ dataset in various scenarios, this article has assessed state-of-the-art instance segmentation techniques and established a standard for disguised instance segmentation. To enhance the performance of these methods even more, a Camouflaged Fusion Learning (CFL) framework was introduced. The results are publicly available on the project page and it provides a valuable resource for further research in camouflaged instance segmentation.

EVPv2 [14], ZoomNeXT [15], and BiRefNet [16] have improved the field of camouflaged object detection and foreground segmentation. EVPv2 was introduced in 2023 and it presents a unified framework for various foreground segmentation tasks like SOD, Defocus Blur Detection, and COD. This paper introduced a novel visual prompting model which is called Explicit Visual Prompting (EVP), and it elevated pre-training and prompt-tuning protocols inspired by NLP. EVP achieved superior performance compared to other fine-tuning methods that are parameter-efficient across multiple datasets and tasks by enforcing tunable parameters concentrating on the unique visual scope of each image. The performance of BiRefNet can be seen in Fig. 5. ZoomNeXT was also introduced in 2023 and addresses the complexity of camouflaged object detection by proposing an effective unified collaborative pyramid network. The model works by zooming strategy to learn discriminative mixed-scale semantics and explores subtle clues between targeted objects and background surroundings. Additionally, ZoomNeXT introduces a simple yet effective regularization called uncertainty awareness loss to support predictions with higher confidence in candidate regions. Its task-friendly framework surpassed the existing state-of-the-art methods in image and video-camouflaged object detection benchmarks.

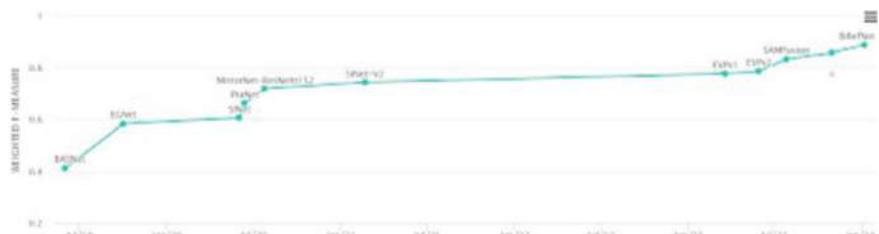
BiRefNet was introduced in 2024 and presents a novel Bilateral Reference framework for enhanced dichotomous image segmentation (DIS). It uses two modules, namely a Localization Module (LM) and a Reconstruction Module (RM) with Bilateral Reference (BiRef) to help in object localization by using global semantic information and for the reconstruction process. The model also introduced additional gradient supervision to improve focus on regions with finer details. BiRefNet shows great performance by outperforming task-specific methods across all benchmarks. Their overall performance can be seen in Fig. 6.

## 5 Methodology

In our methodology, the NC4K dataset serves as the foundation. There are 4121 high-resolution images and each image comes with pre-existing ground truth data in the form of segmented images, highlighting the presence and boundaries of camouflaged



**Fig. 5** Performance of BiRefNet



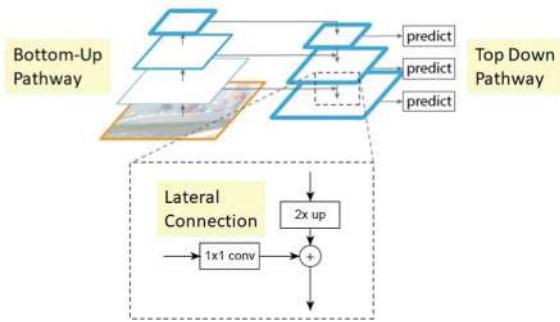
**Fig. 6** Performance of various models on the CAMO dataset

objects. As we already had pre-segmented ground truth, we didn't have to manually segment every image during the training phase.

Our methodology uses the EfficientDet object detection model and specifically the D5 variant known for its balance between accuracy and efficiency. The D5 model is pre-trained on a large image classification dataset and it provides a strong foundation for feature extraction. The NC4K dataset is loaded into the EfficientDet model which allows it to learn robust features that effectively distinguish camouflaged objects from their natural backgrounds.

To know how EfficientDet works, we first have to draw light on Feature Pyramid Network (FPN). It works on the standard idea of executing the algorithm on numerous resolutions of the same image in the hope of catching both small and large-scale

**Fig. 7** Feature pyramid network architecture

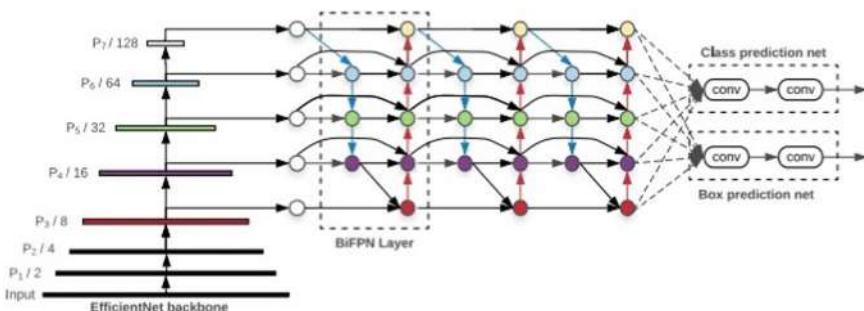


phenomena. Iqbal et al. presented Fig. 7 in their paper [17] that in FPN they use feature maps on different resolutions instead of the different resolution images. In Fig. 7, the traditional backbone of CNN is represented by the bottom-up and the feature fusion at different scales is represented by the top-down. The concept after the lateral connections was to join low-resolution feature maps that are rich in features with less meaningful feature maps with high resolution.

According to Fig. 8 presented by Niu et al. [18] in their paper, the BiFPN acts as the feature network, that continuously applies bottom-up and top-down bi-feature fusion. These mixed features are given to a box network and class to generate bounding box predictions and object class. They are shared at all levels of features equally.

While we used EfficientDet for object detection, we used SAM to further refine the segmentation of camouflaged objects. SAM's zero-shot segmentation model leverages both the image data and the pre-existing ground truth segmentation masks from the NC4K dataset. This combination of SAM and EfficientDet allowed SAM to learn a mapping between image features and object boundaries that have the possibility to enhance the accuracy of the segmentation masks generated by EfficientDet.

We used IoU and Dice coefficient which are standard evaluation metrics which is commonly used in COD and COS tasks. IoU indicates the area of overlap between the predicted and ground truth mask, on the other hand, the Dice coefficient provides



**Fig. 8** Network architecture of EfficientDet

**Table 1** Performance of different models on the NC4K dataset

Model	S-Measure	Weighted F-Measure	MAE	Year
BiRefNet	0.915	0.890	0.0023	2024
ZoomNeXt-PVTv2-B5	0.903	0.863	0.028	2023
ZoomNeXt-PVTv2-B4	0.900	0.865	0.028	2023
ZoomNeXt-ResNet-50	0.874	0.816	0.037	2023
SINetV2-Res2Net-50	0.847	0.770	0.048	2021

a score based on the similarity of intersection and union of these regions. We can quantitatively evaluate the performance of our model by calculating these metrics on the NC4K dataset.

We aim to achieve an accurate and robust system for camouflaged object detection in diverse natural environments by implementing the combined strengths of EfficientDet and SAM. Some comparative results on the NC4K dataset are presented in Table 1.

## 6 Result

The methodology that we have presented achieved a promising result on the NC4K dataset for camouflaged object detection. The Dice coefficient, which is a metric used to calculate the overlap between predicted and ground truth segmentation masks, gave an average score of 87.87%. This high value indicates that our model effectively segments camouflaged objects from their backgrounds.

However, the IoU metric, which is a metric that assesses the intersection of predicted and ground truth masks relative to their union, yielded an average score of 81.18%. While this suggests good object segmentation, it also highlights the potential for improvement in precisely detecting the boundaries of camouflaged objects. The results are visible in Fig. 9.

The high Dice coefficient demonstrates the model's capacity to accurately segment the objects themselves. The moderate IoU score indicates that there's room for improvement in precisely capturing the often-subtle boundaries of camouflaged objects.

## 7 Limitations

Our proposed methodology demonstrates promising results but there are still various limitations that deserve consideration. First is the size of the NC4K dataset. Although the NC4K dataset is large it is not one of the largest datasets. There exist datasets like COD10K that are comparatively larger than the NC4K dataset. Second, there are

**Fig. 9** Results of EfficientDet with SAM on NC4K

```

print("creating sam (model type) and moving it to device")
sam = sam_model_registry[model_type](checkpoint=sam_checkpoint)
sam.to(device=device)
print("creating predictor")
predictor = SamPredictor(sam)
|
datasets = []
datasets.append("NC4K")
# datasets.append("COCO_val2017")

for dataset in datasets:
    perform_all(dataset, predictor, model_type, source_mask)

creating sam default and moving it to device
creating predictor
default oracle NC4K D 50 off
/content/drive/MyDrive/dataset/NC4K/segmentator_oracle/*.bmp
results will be saved at: /content/output/NC4K/oracle/default
100%|██████████| 4121/4121 [3:06:24<00:00, 2.71s/it]SAM alone metrics:
average iou      : 81.18
average dice     : 87.8/

```

chances that the performance of the model might be influenced by the specific characteristics of the NC4K dataset, such as object diversity and image resolutions. We would also like to mention that the current study only focuses on static images which limits its applicability to real-life applications including videos and live footage. Addressing these limitations in future research will contribute to the development of more powerful and adaptable COD systems.

## 8 Future Scope

While our proposed methodology has shown great results in camouflaged object detection, there is still room for improvement and future research. We need to enhance the IoU score. For this, we can fine-tune the model architecture or optimize hyperparameters that might improve the boundary detection accuracy. One more thing that can be done to improve the model is to utilize additional training data or advanced data augmentation techniques to capture the complexities of COD. Experimenting with different models and custom-designed architectures could lead to further improvements in performance. We aim to advance the state-of-the-art in COD and improve detection systems for various real-world applications by exploring these future research directions.

## 9 Conclusion

In conclusion, this paper has presented a novel approach for COD that shows the strengths of EfficientDet and SAM. This approach achieved promising results on the NC4K dataset and demonstrated its effectiveness in detecting camouflaged objects. The Dice coefficient highlighted good segmentation accuracy but the IoU metric

revealed potential for improvement in prediction delineating object boundaries. Further work will focus on refining the model's capability to capture subtle boundaries and explore various strategies like advanced data augmentation and potentially even custom model architectures. Hence, we are determined to push the boundaries of camouflaged object detection by continuously exploring mentioned avenues and benchmarking against the latest advancements.

## References

1. Fan D-P, Ji G-P, Zhou T, Chen G, Fu H, Shen J, Shao L (2020) Pranet: parallel reverse attention network for polyp segmentation. In: International conference on medical image computing and computer-assisted intervention, Springer, pp 263–273
2. Liu W, Shen X, Pun C-M, Cun X (2023) Explicit visual prompting for universal foreground segmentations. arXiv preprint [arXiv:2305.18476](https://arxiv.org/abs/2305.18476)
3. Lv Y, Zhang J, Dai Y, Li A, Liu B, Barnes N, Fan D-P (2021) Simultaneously localize, segment, and rank the camouflaged objects. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 11591–11601
4. Pang Y, Zhao X, Xiang T-Z, Zhang L, Lu H (2024) Zoomnext: a unified collaborative pyramid network for camouflaged object detection. IEEE transactions on pattern analysis and machine intelligence
5. Koonce B, Koonce B (2021) Efficientnet. Convolutional neural networks with swift for Tensorflow: image recognition and dataset categorization, 109–123
6. Le T-N, Nguyen TV, Nie Z, Tran M-T, Sugimoto A (2019) Anabranche network for camouflaged object segmentation. Comput Vis Image Underst 184:45–56
7. Fan D-P, Ji G-P, Cheng M-M, Shao L (2022) Concealed object detection. IEEE transactions on pattern analysis and machine intelligence
8. Tan M, Pang R, Le QV (2020) Efficientdet: scalable and efficient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 10781–10790
9. Fan D-P, Ji G-P, Cheng M-M, Shao L (2021) Concealed object detection. IEEE Trans Pattern Anal Mach Intell 44(10):6024–6042
10. Qin X, Zhang Z, Huang C, Gao C, Dehghan M, Jagersand M (2019) Basnet: boundary-aware salient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 7479–7489
11. Fan D-P, Ji G-P, Cheng M-M, Shao L (2022) Concealed object detection. IEEE Trans Pattern Anal Mach Intell 44(10):6024–6042. <https://doi.org/10.1109/TPAMI.2021.3085766>
12. Zhao J-X, Liu J-J, Fan D-P, Cao Y, Yang J, Cheng M-M (2019) Egnet: edge guidance network for salient object detection. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 8779–8788
13. Le T-N, Cao Y, Nguyen T-C, Le M-Q, Nguyen K-D, Do T-T, Tran M-T, Nguyen TV (2021) Camouflaged instance segmentation in-the-wild: dataset, method, and benchmark suite. IEEE Trans Image Process 31:287–300
14. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo W-Y, et al (2023) Segment anything. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 4015–4026
15. Le T-N, Cao Y, Nguyen T-C, Le M-Q, Nguyen K-D, Do T-T, Tran M-T, Nguyen TV (2022) Camouflaged instance segmentation in-the-wild: dataset, method, and benchmark suite. IEEE Trans Image Process 31:287–300. <https://doi.org/10.1109/TIP.2021.3130490>
16. Zheng P, Gao D, Fan D-P, Liu L, Laaksonen J, Ouyang W, Sebe N (2024) Bilateral reference for high-resolution dichotomous image segmentation. arXiv preprint [arXiv:2401.03407](https://arxiv.org/abs/2401.03407)

17. Iqbal S, Qureshi AN, Li J, Mahmood T (2023) On the analyses of medical images using traditional machine learning techniques and convolutional neural networks. *Arch Comput Methods Eng* 30(5):3173–3233
18. Niu S, Zhou X, Zhou D, Yang Z, Liang H, Su H (2023) Fault detection in power distribution networks based on comprehensive-yolov5. *Sensors* 23(14):6410

# Blockchain for Cloud/Edge/Fog Computing: A Review



Soukeina Zouaidi

**Abstract** This review paper addresses the use of Blockchain technology in cloud, edge and fog computing environments. It introduces a novel taxonomy of this use regarding the security aspects of the fundamental features of these computing paradigms: data storage and data processing. The study has facilitated the identification of key findings and outlined perspectives.

**Keywords** Cloud computing · Edge computing · Fog computing · Security · Data storage · Data processing

## 1 Introduction

Cloud, edge and fog computing are paradigms that revolutionize the distribution of computing resources. Cloud computing provides centralized services over the internet. Meanwhile, edge and fog computing distribute computing resources closer to endpoint devices in order to reduce the latency.

The use of Blockchain technology for cloud/edge/fog computing represents a paradigm shift to address several challenges inherent to these environments. This use was addressed in previous studies. For example, [1] addressed the employment of Blockchain for the reengineering of cloud data centers. Yang et al. [2] addressed the incorporation of Blockchain within edge computing paradigm. Baniata and Kertesz [3] addressed the integration of Blockchain with fog computing according: the used algorithms, Blockchain functions and the location of the Blockchain in the fog network. However, these works did not address this integration regarding the main features of these paradigms. For this reason, we propose this review paper to address the use of Blockchain for cloud, edge and fog computing paradigms while focusing on the security of features provided by these paradigms, namely data storage and data processing. The main contributions of this review paper are threefold: First, we

---

S. Zouaidi (✉)  
ESPRIT School of Engineering, Tunis, Tunisia  
e-mail: [soukeina.zouaidi@esprit.tn](mailto:soukeina.zouaidi@esprit.tn)

provide a new taxonomy of Blockchain utilization. Second, we present our analysis of the reviewed literature. Finally, we identify the future research directions.

The remainder of this paper is organized as follows: Sect. 2 introduces the review context. Section 3 presents the new taxonomy of the literature. Finally, Sect. 4 gives the main findings and outlines the perspectives.

## 2 Review Context

This section presents the review context. The first subsection introduces the Blockchain technology, and the second subsection introduces the computing paradigms (cloud, edge and fog computing).

### 2.1 *Blockchain*

Blockchain technology is a distributed architecture emerged from the use of digital ciphered currency (e.g., Bitcoin) [4]. This technology revolutionizes the trust of decentralized data management. Indeed, it eliminates the requirement for a centralized trusted authority by enabling peer to peer interactions. Blockchain is represented as a linked data structure where data blocks are securely interconnected in a chronological and cryptographic manner, ensuring resistance to tampering. The use of Blockchain for cloud, edge and fog computing has revolutionized data management and processing.

### 2.2 *Computing Paradigms*

#### 2.2.1 **Cloud Computing**

Cloud computing offers virtualized resources over internet [5]. These resources are used remotely without the requirement for owning or managing physical infrastructure. This paradigm offers several advantages, including scalability and accessibility from anywhere. It has revolutionized the data storage and processing without the complexities and expenses associated with traditional IT infrastructures.

#### 2.2.2 **Edge Computing**

Edge computing utilizes the storage and processing resources of various edge devices such as IoT devices ...[6] acting as an intermediate layer between endpoint devices and the cloud. These edge devices reduce the workload on data centers by handling

some remote requests locally. This decreases latency and enables real-time processing for certain applications.

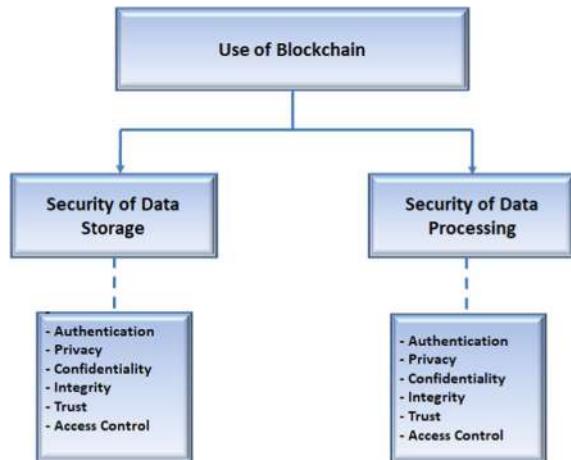
### 2.2.3 Fog Computing

Fog computing extends the reach of cloud computing from its centralized core to the network edge. Indeed, it enables computation closer to end-user devices by leveraging local computing resources on the edge [7]. This approach reduces the data transfer to the central cloud, thereby ensuring low latency.

## 3 Use of Blockchain

This section introduces the use of Blockchain for the computing paradigms: cloud, edge and fog computing. We have reviewed the use of this technology based on the security of the main features provided by these paradigms which are: data storage and data processing. To structure our review, we have classified this use into two main topics: the first is security of data storage and the second is security of data processing as illustrated in Fig. 1.

**Fig. 1** Taxonomy of blockchain use for cloud/edge/fog computing



### 3.1 Security of Data Storage

In the following, we review some works which have used the Blockchain as a tool to secure the data storage.

ChainFS [8] is a Blockchain-based middleware that ensures the security of cloud storage. This middleware prevents forking attacks. To achieve this, the data files are stored in the cloud, while the key management and file manipulation logging are implemented into the Blockchain. The performance evaluation of this system demonstrates low overhead.

Yugala [9] is a Blockchain-based architecture for cloud storage. It ensures file confidentiality and integrity. In addition, it eliminates the centralized data redundancy. The performance evaluation of Yugala shows its efficiency for handling large files size while respecting the Blockchain load.

The security of cloud files sharing was addressed in [10]. The proposed solution is based on Blockchain and attribute-based encryption. Blockchain is used to establish a smart contract for access management between data holder and utilizers. Attribute-based encryption is used in response transactions to enforce the access policy to the file while also maintaining user anonymity. The evaluation results show the effectiveness of this scheme in terms of scalability.

Blockchain was used in [11] to ensure trusted collaborative edge storage. The offloading decision is based on the reputation concept. Indeed, the selection of participants is based on their reputation that is assessed after each task completion and stored on the Blockchain for subsequent assessments, through a consensus mechanism.

A hierarchical access control mechanism was proposed in [12]. This mechanism is based on the classification of users and data to achieve dynamicity and credibility. Indeed, a security level is assigned for each data resource and each user is assigned to a credibility group. The different security levels and credibility groups are stored into the Blockchain. The performance evaluation of this mechanism shows efficiency regarding delay, throughput and processor usage.

A Blockchain-based audit scheme for cloud storage was introduced in [13]. This scheme aims to achieve data privacy by leveraging the computing resources of edge servers. Indeed, the edge computing servers fulfills the proof of verification. The performance assessment of this proposal shows efficiency in terms of the computational cost of completeness audit.

The security of the distributed data storage was addressed in [14]. The proposed solution is based on Blockchain technology and fog computing model. To achieve security, authentication and access control processes are implemented. In addition, the Hybrid Encryption Algorithm (HEA) is used to improve security by merging various cryptographic mechanisms for strong prevention from illegitimate access.

FogChain [15] is a healthcare framework. It aims to ensure the privacy of collected medical records. It is designed based on IoT, fog computing and Blockchain. The performance evaluation of FogChain shows that the use of an intermediary fog layer near to the edge improves the performance in terms of response time in comparison with Cloud-like Blockchain infrastructures.

The security of medical records was addressed also in [16]. The proposed solution aims to alleviate the load of IoT devices by outsourcing the computation to fog nodes. To achieve the security requirements, it implements an access control mechanism. The patients' privacy is guaranteed by storing the operations on a Blockchain. Additionally, the Blockchain is augmented with a multiauthority attribute-based encryption scheme to ensure data confidentiality. This solution shows its efficiency in terms of reliability and practicality.

### 3.2 *Security of Data Processing*

In the following, we review some works which have used the Blockchain as a tool to secure data processing.

Verifiable computation using smart contracts over a cryptocurrency Blockchain was addressed in [17]. Ensuring accurate computation results is based on delegating computation to two cloud servers and requiring at least one of the two servers is honest. The smart contract serves as a Trusted Third Party (TTP) to facilitate interactions between parties and manage fund transfer between them.

Blockchain was used in [18] to build a secure mobile adhoc cloud architecture. The main design objectives of this architecture are: anonymity, auditability, persistency and decentralization. This architecture includes three types of participants, which are job publisher, helper and job handler. The job publisher initiates job by establishing a new smart contract. Helpers leverage their extra computing resources to execute the published tasks and, in return, receive incentives. The job handler, represented by the whole Blockchain, provides the running framework.

Blockchain was used also in [19] to design a secure job scheduling model for cloud computing. This model aims to prevent threats like job injection, machine defect or generation of wrong schedule.

D2D-ECN [20] is an edge computing framework which aims to improve the real-time running of IoT applications. This framework is based on Blockchain technology to secure resource trading and task assignment. The reliability of this framework is achieved through a consensus mechanism driven by reputation. The effectiveness of computation offloading in D2D-ECN was validated through feasibility analysis and numerical results.

Blockchain was used in [21] to ensure trust edge collaboration while taking into account also user privacy. The use of this technology ensures a constant and immutable records of job execution. The selection of the computing node is based on a reputation score which is calculated based on execution efficiency. The performance evaluation of this solution shows higher efficiency of edge resources use.

A content caching and computation model for edge computing was introduced in [22]. This model leverages Blockchain technology to ensure the integrity of cached content and the authentication of users. The performance assessment of this proposed model demonstrates both cache efficiency and delay minimization. However, further exploration is needed to address energy efficiency concerns in this work.

Blockchain technology was used in [23] to achieve the security of cooperative computing strategy. To enhance the security level, the fog environment is divided into clusters. An access control list is associated to each cluster. The results of access decisions are stored in the Blockchain. The simulation results show that the cooperative strategy reduces the computing time of a block hash in comparison to the non-cooperative one.

Blockchain technology was used also in [24] to ensure trust cooperative offloading between fog nodes of multiple applications. In addition, a public key infrastructure was used in to authenticate nodes with lower reputation. The security analysis of this approach shows its ability to prevent unverified nodes from undertaking offloading.

The security of computing resource sharing was addressed in [25] by the use of Blockchain. The proposed solution is based on a reward mechanism secured by the Blockchain. This mechanism incentives participation for sharing resources by granting fixed credits via the Blockchain system.

## 4 Main Findings and Perspectives

This section summarizes this review and gives some perspectives.

### 4.1 Main Findings

This paper reviews the use of Blockchain for the computing paradigms: cloud, edge and fog computing. When we have investigated the purpose of this integration, we have proposed a new classification of the use of this technology for these paradigms. This classification is based on the security objective related to the two main features of these paradigms: data storage and data processing. Therefore, we have outlined this classification as the security of data storage and the security of data processing. When investigating the security purpose of this integration, we have explored the security services fulfilled by Blockchain, which are: authentication, privacy, confidentiality, integrity, trust and access control.

The reviewed works are summarized in two tables, referenced as Tables 1 and 2. The first table provides a summary of the reviewed works based on the main use of Blockchain, namely: security of data storage and security of data processing. Meanwhile, the second table outlines the security services addressed in each work, including authentication, privacy, confidentiality, integrity, trust and access control. In addition, the limitations associated to each addressed security service are summarized in Table 3.

The authentication service was addressed in [14] to authenticate users to access stored data. This service was addressed also in [24] to authenticate the node to which the task will be offloaded. The user authentication was addressed also in [22] to

**Table 1** Taxonomy of the reviewed literature based on blockchain use

Paradigm	Paper	Security of data storage	Security of data processing
Cloud computing	[8]	✓	
	[9]	✓	
	[10]	✓	
	[17]		✓
	[18]		✓
	[19]		✓
Edge computing	[11]	✓	
	[12]	✓	
	[13]	✓	
	[20]		✓
	[21]		✓
	[22]		✓
Fog computing	[14]	✓	
	[15]	✓	
	[16]	✓	
	[23]		✓
	[24]		✓
	[25]		✓

prevent fake content requests. However, it should be noted that the scalability of this service was not addressed.

The data privacy was addressed in [13, 15, 16] while user privacy was addressed in [18, 21]. However, it should be noted that the location privacy and conditional privacy were not addressed. Indeed, in particular, the protection of user privacy should be conditional in order to ensure the tracing of any malicious behavior.

The data confidentiality was addressed in [9, 16]. However, it should be noted that the confidentiality of computation was not addressed. Indeed, some computational results could be saved in cloud servers or edge nodes. Thus, these results should be protected from being eavesdropped.

The data integrity was addressed in [8, 9] while the integrity of computation was addressed in [17, 22]. In addition, the integrity of task scheduling was addressed in [19]. However, it should be noted that the computational and communication overhead related to the integrity protection were not addressed.

The trust of data storage was addressed in [11] while the trust of the computing entity was addressed in [20, 21, 24, 25]. However, it should be noted that the efficiency of trust model in terms of precision and accuracy was not addressed.

The access control for stored data was addressed in [10, 12, 14] while the access control for computation was addressed in [23]. However, it should be noted that the granularity of this service was not addressed.

**Table 2** Security services for each reviewed work

Paradigm	Paper	Authentication	Privacy	Confidentiality	Integrity	Trust	Access control
Cloud computing	[8]				✓		
	[9]			✓	✓		
	[10]						✓
	[17]				✓		
	[18]		✓				
	[19]				✓		
Edge computing	[11]					✓	
	[12]						✓
	[13]		✓				
	[20]					✓	
	[21]		✓			✓	
	[22]	✓			✓		
Fog computing	[14]	✓					✓
	[15]		✓				
	[16]		✓	✓			✓
	[23]						✓
	[24]	✓				✓	
	[25]					✓	

**Table 3** Limitations for each security service

Security service	Limitations
Authentication	Scalability was not addressed
Privacy	Location privacy and conditional privacy were not addressed
Confidentiality	Confidentiality of computational results was not addressed
Integrity	Computational and communication overhead were not addressed
Trust	Efficiency of trust model was not addressed
Access control	Granularity was not addressed

## 4.2 Perspectives

Besides the above-indicated limitations of the existing literature, we further propose the following research directions.

- **Network Resilience:** The dynamic nature of cloud/edge/fog computing (e.g., joining and leaving of nodes or failure of cloud server) leads to potential disruptions in data consistency and availability. Therefore, the Blockchain technology should integrate mechanisms to ensure the resilience to changes.
- **Network Security:** Data is collected from the cloud/edge/fog devices through a dedicated network and it can be used for further processing. However, after the data is collected from the network devices, it might be intercepted or falsified during the propagation process. Therefore, the integration of Blockchain should incorporate monitoring mechanisms.
- **Real-Time Processing:** Given the distributed nature of cloud/edge/fog computing, latency management is critical to meet real-time processing requirement for applications like IoT, autonomous vehicles and smart cities. The latency due to consensus protocols, may not achieve this requirement. Thus, the integration of Blockchain should take into consideration this constraint.
- **Scalability:** The Blockchain system often rely on a single ledger and a fixed consensus mechanism. When the network needs to manage a large number of users simultaneously, this can affect the performance. The integration solution of Blockchain should scale effectively.
- **Use of Artificial Intelligence:** The Blockchain is used to establish a tamper-proof record of all transactions in a cloud/edge/fog computing network. This solution could be enhanced by the use of artificial intelligence models to detect anomalies (e.g., malicious activity) and automatically take corrective measures.

## 5 Conclusion

In this paper, we reviewed the use of Blockchain for the computing paradigms: cloud, edge and fog computing. During our investigation of this issue, we proposed a new taxonomy of this issue into tow main topics: security of data storage and security of data processing. Finally, we gave an analysis of the reviewed literature along with perspectives.

## References

1. Gai K, Raymond Choo K-K, Liehuang Z (2018) Blockchain-enabled reengineering of cloud datacenters. *IEEE Cloud Comput* 5(6):21–25
2. Yang R, Yu F-R, Si P, Yang Z, Zhang Y (2019) Integrated blockchain and edge computing systems: a survey, some research issues and challenges. *IEEE Commun Surv Tutor* 21(2):1508–1532
3. Baniata H, Kertesz A (2020) A survey on blockchain-fog integration approaches. *IEEE Access* 8:102657–102668
4. Nakamoto S (2008) Bitcoin: a peer-to-peer electronic cash system
5. Vaquero LM, Rodero-Merino L, Caceres J, Lindner M (2008) A break in the clouds: towards a cloud definition. *ACM SIGCOMM Comput Commun Rev* 39(1):50–55
6. Shi W, Cao J, Zhang Q, Li Y, Xu L (2016) Edge computing: vision and challenges. *J IEEE IoT* 3(5):637–646
7. Bonomi F, Milito R, Zhu J, Addepalli S (2012) Fog computing and its role in the internet of things. In: Proceedings of the ACM workshop on mobile cloud computing, Helsinki, 17 August
8. Tang Y, Zou Q, Chen J, Li K, Kamhoua CA, Kwiat K, Njilla L (2018) ChainFS: blockchain-secured cloud storage. In: Proceedings of the IEEE international conference on cloud computing, San Francisco, 02–07 July
9. Gochhayat SP, Bandara E, Shetty S, Foyti P (2019) Yugala: blockchain based encrypted cloud storage for IoT data. In: Proceedings of the IEEE international conference on blockchain, Atlanta, 14–17 July
10. Almasian M, Shafieinejad A (2024) Secure cloud file sharing scheme using blockchain and attribute-based encryption. *J Comput Stand Interfaces* 87
11. Yuan L, He Q, Chen F, Zhang J, Qi L, Xu X, Xiang Y, Yang Y (2022) CSEdge: enabling collaborative edge storage for multi-access edge computing based on blockchain. *IEEE Trans Parallel Distrib Syst* 33(8):1873–1887
12. Hou Y, Liu W, Lin H, Wang X (2020) Multi-layer access control mechanism based on blockchain for mobile edge computing. In: Proceedings of the IEEE international conference on parallel distributed processing with applications, big data & cloud computing, sustainable computing communications, social computing networking, exeter, 17–19 December
13. Wang J, Wang S, Wang L, Shao W, Xu S, Zhang S (2022) A blockchain and edge computing based public audit scheme for cloud storage. In: Proceedings of the Chinese control conference, Hefei, 25–27 July
14. Agrawal R, Singhal S, Sharma A (2024) Blockchain and fog computing model for secure data access control mechanisms for distributed data storage and authentication using hybrid encryption algorithm. *Cluster Comput*
15. Mayer HA, Rodrigues VF, Costa CA, Rosa Righi R, Roehrs A, Antunes RS (2021) Fogchain: a fog computing architecture integrating blockchain and internet of things for personal health records. *IEEE Access* 9:122723–122737
16. Li J, Li D, Zhang X (2023) A secure blockchain-assisted access control scheme for smart healthcare system in fog computing. *IEEE IoT J* 10(18):15980–15989
17. Avizheh S, Nabi M, Safavi-Naini R, Venkateswarlu KM (2019) Verifiable computation using smart contracts. In: Proceedings of the ACM conference on cloud computing security workshop, London, 11 November
18. Jiao Z, Zhang B, Zhang L, Liu M, Gong W, Li C (2020) A blockchain based computing architecture for mobile ad-hoc cloud. *J IEEE Netw* 34(4):140–149
19. Wilczynski A, Kolodziej J, Grzonka D (2021) Security aspects in blockchain-based scheduling in mobile multi-cloud computing. In: Proceedings of the IEEE/ACM international symposium on cluster, cloud and internet computing, Melbourne, 10–13 May
20. Qiao G, Leng S, Chai H, Asadi A, Zhang Y (2019) Blockchain empowered resource trading in mobile edge computing and networks. In: Proceedings of the IEEE international conference on communication, Shanghai, 20–24 May

21. Gao Q, Xiao J, Cao Y, Deng S, Ouyang C, Feng Z (2023) Blockchain-based collaborative edge computing: efficiency, incentive and trust. *J Cloud Comput* 12
22. Bozkaya-Aras E (2024) Blockchain-based secure content caching and computation for edge computing. *IEEE Access* 12:47619–47629
23. Wu D, Ansari N (2020) A cooperative computing strategy for blockchain-secured fog computing. *IEEE IoT J* 7(7):6603–6609
24. Roshan R, Matam R, Mukherjee M, Lloret J, Tripathy S (2020) A secure taskoffloading framework for cooperative fog computing environment. In: Proceedings of the IEEE global communications conference, Taipei, 07–11 December
25. Rani S, Gupta D, Herencsar N, Srivastava G (2023) Blockchain-enabled cooperative computing strategy for resource sharing in fog networks. *J IoT* 21

# Secure Data Communication: Implementation and Performance Evaluation of Aggregate Key Encryption



Nagabhyru Nikitha, Kogatam Vijayasree, Nara Bhavyasree, and D. Radha

**Abstract** Encryption is essential for securing sensitive information from unauthorized access, ensuring confidentiality and privacy. It safeguards data integrity, preventing tampering or alteration during transmission or storage. Encryption also facilitates authentication, verifying the identities of users or systems, and enabling trusted communication. It helps organizations comply with regulatory requirements and maintain control over their data, especially in cloud environments. Ultimately, encryption is indispensable for building trust in digital interactions and protecting against cyber threats. This study presents an investigation into the implementation and analysis of encryption of data using aggregate keys, comparing it to traditional encryption methods such as RSA and AES. Aggregate key encryption offers a novel approach to managing encryption keys, aggregating multiple keys into a single entity for enhanced efficiency and scalability. The research involves the development of algorithms and protocols for generating, distributing, and utilizing aggregate keys securely. The algorithm is designed in such a way that it encrypts user access control encryption and normal encryption like AES, and RSA. Additionally, a comprehensive analysis is conducted to assess the security, performance, and practicality of aggregate key encryption in comparison to RSA and AES. The study aims to provide insights into the effectiveness and efficiency of aggregate key encryption, contributing to advancements in data security and encryption practices.

**Keywords** Algorithm · Decryption · Encryption

---

Supported by organization x.

N. Nikitha · K. Vijayasree (✉) · N. Bhavyasree · D. Radha

Department of Computer Science Engineering, Amrita School of Computing, Bengaluru Amrita

Vishwa Vidyapeetham, Bengaluru, India

e-mail: [vijayasreekogatam2002@gmail.com](mailto:vijayasreekogatam2002@gmail.com)

D. Radha

e-mail: [d.radha@blr.amrita.edu](mailto:d.radha@blr.amrita.edu)

## 1 Introduction

In the realm of data security, encryption plays a pivotal role in safeguarding sensitive information from unauthorized access and breaches. Traditional encryption methods typically involve the use of individual keys for securing data, which can become cumbersome to manage, especially in scenarios involving large-scale data storage and transmission. To address this challenge, the concept of aggregate key encryption has emerged as a promising solution.

Aggregate key encryption involves the aggregation of multiple keys into a single entity, known as an aggregate key, which is used to encrypt and decrypt data. This approach offers several advantages, including reduced key management overhead, enhanced scalability, and improved efficiency in key distribution and access control.

The implementation of encryption using aggregate keys entails the development of algorithms and protocols to generate, distribute, and utilize aggregate keys securely. This implementation process involves various cryptographic techniques, such as symmetric and asymmetric encryption, hash functions, and key management schemes.

Furthermore, the analysis of encryption using aggregate keys involves assessing the security, performance, and practicality of the proposed encryption scheme. Security analysis includes evaluating the resistance of the encryption scheme against various cryptographic attacks, such as brute force attacks, differential cryptanalysis, and chosen plaintext attacks.

Performance analysis focuses on quantifying the computational overhead, memory requirements, and communication overhead associated with the encryption and decryption processes using aggregate keys. This analysis helps in understanding the efficiency and scalability of the encryption scheme, particularly in real-world deployment scenarios.

Moreover, the practicality analysis considers factors such as ease of implementation, compatibility with existing systems, and compliance with regulatory requirements. It involves assessing the feasibility of integrating the encryption scheme into different applications and environments while ensuring usability and maintainability.

Overall, the implementation and analysis of encryption using aggregate keys represent a significant advancement in the field of data security, offering a viable solution for addressing key management challenges and enhancing the confidentiality and integrity of sensitive information in various domains, including cloud computing, Internet of Things (IoT), and distributed systems. This study aims to explore and evaluate the effectiveness and efficiency of aggregate key encryption, contributing to the advancement of secure data communication and storage practices in modern digital ecosystems.

## 2 Literature Survey

The concept of key-aggregate encryption (KASE) has emerged as a significant advancement in the realm of secure and efficient group data sharing through cloud storage. KASE addresses the critical need for enabling data owners to share encrypted information with multiple users within a group, while simultaneously empowering each user to conduct searches over the shared data without compromising the actual content to the cloud storage provider. However, the existing KASE schemes grapple with the challenge of dynamic group management. Specifically, the need to regenerate the entire joint decryption key during user revocation or update poses considerable computational burden, particularly for large groups undergoing frequent changes. This limitation underscores the necessity for further exploration and refinement of KASE [3] schemes to enhance their scalability and efficiency in handling dynamic group scenarios. To delve deeper into the literature surrounding KASE, researchers have focused on various aspects, such as the cryptographic techniques employed for secure data sharing, the computational complexities associated with dynamic group management, and the trade-offs between security and efficiency. Additionally, studies have explored potential enhancements to KASE schemes, seeking novel approaches to streamline key regeneration processes and minimize computational overhead. Analyzing the existing body of literature on KASE not only highlights its potential benefits but also sheds light on the ongoing challenges that researchers are actively addressing. Further investigation into the latest developments in this field is crucial for refining KASE schemes and ensuring their practical applicability in real-world scenarios of group data sharing via cloud storage.

Key-aggregate-based access control encryption (KA-ACE) presents a novel approach to achieving fine-grained and efficient access control for cloud data sharing. By leveraging key-aggregate cryptosystems, the scheme aims to enhance both the security and scalability of access control mechanisms for users and data in cloud environments. Unlike traditional access control encryption schemes, KA-ACE offers improved efficiency. However, the scheme faces a challenge in terms of scalability due to the necessity of regenerating the entire joint decryption key during user revocations or updates. The literature surrounding KA-ACE [4] explores various cryptographic techniques, efficiency optimizations, and security considerations related to its implementation. Researchers have delved into the cryptographic foundations of KA-ACE, investigating the underlying principles that contribute to its fine-grained access control capabilities. The literature review emphasizes the significance of key-aggregate cryptosystems in achieving the desired level of security while maintaining efficiency. Scholars have also focused on analyzing the scalability issues arising from the need to regenerate the joint decryption key, proposing potential solutions to mitigate computational overhead and enhance the scheme's applicability in scenarios with dynamic user access requirements.

Comparative analyses between KA-ACE and traditional access control encryption schemes [5] have been a focal point in the literature. Researchers have examined

the trade-offs between security and efficiency, highlighting the advantages of KA-ACE in terms of providing flexible and scalable access control. Investigations into the scheme's real-world applicability and performance in cloud data-sharing environments have been carried out, offering insights into potential areas of improvement and further development. Furthermore, literature on KA-ACE often explores extensions or modifications to the original scheme to address identified limitations. This may involve refining key management strategies, optimizing cryptographic operations, or proposing alternative approaches to dynamic access control in cloud environments. Overall, the literature review on KA-ACE provides a comprehensive understanding of the scheme's strengths, limitations, and ongoing research efforts aimed at advancing its capabilities for flexible and secure cloud data sharing.

Hybrid cryptography for secure file storage has emerged as a prominent paradigm for ensuring the confidentiality and integrity of sensitive data files. The approach combines the advantages of both symmetric and public-key cryptography [6], providing a robust solution that is particularly well-suited for securing large volumes of data stored in cloud environments. The literature surrounding hybrid cryptography delves into various aspects, including its foundational principles, key management challenges in cloud settings, and potential enhancements to address these challenges. Researchers have extensively explored the cryptographic underpinnings of hybrid schemes, investigating the synergies between symmetric and public-key cryptography. The literature review highlights the strengths of symmetric algorithms in terms of efficiency and speed for bulk data encryption, complemented by the asymmetric algorithms' ability to securely exchange keys and facilitate secure communication channels. Understanding the balance and integration of these cryptographic techniques is crucial for optimizing the overall security of file storage systems.

The study of AES and RSA algorithms based on Graphics Processing Units (GPUs) [8] represents a significant exploration into leveraging the parallel processing capabilities of GPUs to enhance the performance of encryption operations. The literature survey on this topic provides insights into the feasibility, challenges, and advancements in GPU-based implementations of AES and RSA.

The literature emphasizes the demonstrated feasibility of employing GPUs for accelerating the execution of AES and RSA encryption algorithms. Researchers have extensively investigated the parallelizability of these cryptographic algorithms and proposed efficient GPU-based implementations [9]. The survey highlights the advantages of GPUs in handling parallel tasks, making them particularly well-suited for the data intensive nature of cryptographic operations.

### 3 Methodology

#### 3.1 Key-Aggregate-Based Authentication (KAU)

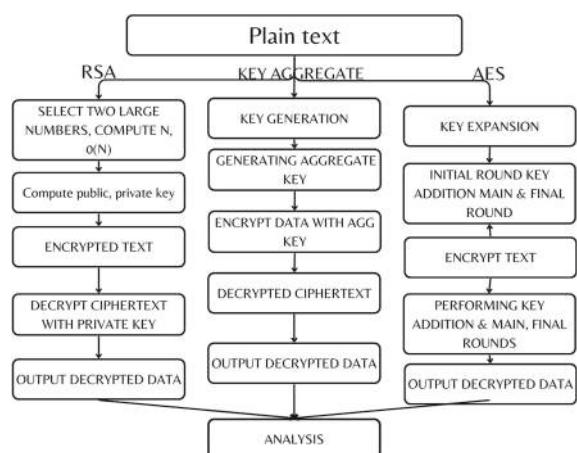
Context: KAA, as a method of authentication, involves users providing a combination of private keys to access a secure system. This approach is chosen for its adaptability in scenarios where multiple users need to collaborate securely, and fine-grained access control is paramount. The architecture draws inspiration from key-aggregate encryption, ensuring efficient and secure management of cryptographic keys.

Security: While key-aggregate authentication offers a robust mechanism for secure access, certain vulnerabilities need consideration. The cryptographic strength of the system relies on the randomness of generated keys and the correctness of the implementation. The method may be susceptible to attacks such as key compromise, unauthorized aggregation, or collusion among users. The security of the system is contingent on addressing these potential threats (Fig. 1).

Advantages: The advantages of KAA lie in its ability to provide a streamlined and secure method for authentication, especially in collaborative environments. The use of aggregate keys reduces the key management overhead, enhancing overall system efficiency. Additionally, KAA allows for dynamic access control, enabling administrators to adapt user privileges dynamically.

Weaknesses: Despite its advantages, KAA is not immune to certain weaknesses. The security of the system heavily relies on the correctness of the key-aggregate algorithm and the randomness of generated keys. Inadequate implementation or vulnerabilities in the key-aggregate mechanism may compromise the overall security of the authentication process.

**Fig. 1** Model diagram



**Enhancements:** To bolster the security of the key-aggregate authentication system, several enhancements can be considered:

**Advanced Key Generation:** Implementing a more sophisticated key generation process that incorporates secure random number generation and ensures the uniqueness of keys. **Secure Key Storage:** Employing secure key storage mechanisms, such as hardware security modules or trusted execution environments, to protect cryptographic keys from unauthorized access. **Dynamic Access Control Policies:** Enhancing the adaptability of access control policies to accommodate changing user roles and permissions dynamically. **Cryptographic Auditing:** Implementing robust cryptographic auditing mechanisms to detect and respond to potential security breaches or unauthorized activities. Incorporating these enhancements can fortify the key-aggregate-based authentication method, ensuring a resilient and secure approach to user access control in collaborative environments.

### ***Algorithm***

#### **Set Up**

Choose a secure prime number p using  $p = \text{getPrime}(128)$ .

- Select a generator G.

#### **Key Generation**

- Generate a list of users = Alice, Bob, Charlie.
- Generate a private key  $\text{private}@\text{key}$  for each user as a random values between 1 and  $p-1$ .
- Calculate public key  $\text{public-keys}$  for each user as  $\text{public keys}[user] = (G \text{ pow}(\text{private-key}[user])) \text{ mod } p$ .

#### **Encryption**

- Read a file specified by file-path and convert its content to an integer message.
- Calculate the aggregate key aggregate-key by summing the private keys of users in the subset.
- Generate a random number r between 1 and  $p-1$ .
- Compute:  $C1 = r * p$   $C2 = [\text{message} + (\text{aggregate-key} \text{ mod } G)]$   $C3 = 1$ . For each user in subset update  $C3 = (C3 * (\text{public-key}[user] \text{ pow}(r) \text{ mod } G))$ .
- Store C1, C2, C3 values as base64 encoded strings in JSON format and write them to ‘encrypted-file.txt’.

#### **Decryption**

- Read the contents of ‘encrypted-file.txt’.
- Decode the base64 strings back to integers (C1, C2, C3).
- For each user in the subset: Calculate  $Ki = (Ci \text{ pow } \text{private-key}[user] \text{ mod } G)$ . Calculate  $Di = (C3 * \text{inverse}(Ki, G) \text{ mod } G)$ . Store Di values in a list D-values.
- Compute the shared key K by multiplying all Di values modulo G.

- Calculate the decrypted messages as decrypted-message = $(C_2 + (K \bmod G))$ .
- Write the decrypted message to ‘decrypted-file.txt’

### ***3.2 RSA-Based Authentication***

This authentication method relies on users providing a combination of a public key (username) and a private key (password) to access a system. RSA is widely used for its robust security features and versatility in securing digital communication.

**Security:** While RSA-based authentication offers strong security, it can be susceptible to certain attacks, including brute force, chosen ciphertext attacks, and side-channel attacks. The security relies heavily on the key length chosen, and smaller key sizes may be vulnerable to attacks.

**Advantages:**

**Strong Security:** RSA is based on the mathematical complexity of factoring large numbers, providing a high level of security. **Versatility:** RSA can be used for both encryption and digital signatures, making it suitable for various security applications.

**Weaknesses:**

**Key Length Considerations:** The security of RSA is directly tied to the key length, and shorter key lengths may be vulnerable to attacks. **Resource Intensive:** RSA operations, especially key generation, can be computationally intensive. **Enhancements:**

**Key Length Selection:** Choose appropriate key lengths based on current security standards to resist attacks. **Secure Key Storage:** Implement secure storage mechanisms for private keys to prevent unauthorized access. **Periodic Key Renewal:** Regularly update and renew RSA keys to mitigate the risk of long-term vulnerabilities. **Multi-factor Authentication (MFA):** Combine RSA-based authentication with additional factors (such as passwords or biometrics) to enhance overall security.

### ***3.3 Advanced Encryption Standard (AES) Encryption***

**Context:** AES encryption is a widely utilized method for securing sensitive data by transforming it into an unreadable format, typically through the use of a secret key. It is employed in various applications, such as securing communications, file encryption, and database protection.

**Security:** While AES is renowned for its robust security, vulnerabilities can still arise, especially in the way it's implemented or managed. Potential risks include key compromise, side-channel attacks, and issues with key storage.

**Advantages:** The popularity and widespread adoption of AES stem from its proven security track record and efficiency. Its standardization and acceptance across industries contribute to its reliability.

**Weaknesses:** Despite its strength, AES is not immune to all forms of attacks. Weaknesses may arise from poor key management, implementation flaws, or potential vulnerabilities in specific modes of operation.

**Enhancements:** To bolster the security of AES encryption, several enhancements and best practices can be implemented:

#### **Key Management:**

Robust key management practices are crucial. Regularly update and rotate encryption keys to mitigate the risk of compromise. Implement secure key storage mechanisms to prevent unauthorized access to encryption keys.

#### **Mode of Operation:**

Choose an appropriate mode of operation (e.g., CBC, GCM) based on the specific security requirements of the application. Be aware of potential vulnerabilities associated with specific modes and use appropriate countermeasures.

#### **Implementation of Best Practices:**

Ensure that AES is implemented correctly and follows recommended best practices. Regularly update cryptographic libraries and tools to patch any discovered vulnerabilities.

#### **Secure Communication:**

When using AES in communication protocols, ensure the overall security of the communication channel (e.g., TLS/SSL). Implement additional security measures, such as secure key exchange protocols.

#### **Key Length and Algorithm Configuration:**

Choose an appropriate key length (e.g., 128-bit, 256-bit) based on the desired level of security. Stay informed about advancements in cryptographic research to adapt to changing security requirements. By incorporating these enhancements and best practices, the security posture of AES encryption can be significantly strengthened, addressing potential weaknesses and ensuring a higher level of protection for sensitive data.

### ***3.4 File Encryption Process***

Key Generation: Generate a unique AES key for each file. RSA Encryption: Encrypt the AES key using the recipient's RSA public key. This step ensures secure key exchange. Key-Aggregate Operation: Optionally, perform a key-aggregate operation to combine multiple keys into a single key for more efficient management. AES Encryption: Encrypt the actual file or image data using the generated AES key. Use a secure mode like CBC for AES encryption to provide confidentiality and integrity. Store Encrypted Data: Save the encrypted file or image data along with the RSA-encrypted AES key.

### ***3.5 File Decryption Process***

RSA Decryption: Use the recipient's RSA private key to decrypt the RSA-encrypted AES key. AES Decryption: Decrypt the file or image data using the decrypted AES key. Utilize the same AES mode used during encryption. Retrieve Original Data: Retrieve the decrypted file or image data, which is now in its original form. Security Considerations: Key Management: Implement secure key management practices to protect both symmetric and asymmetric keys. Authentication: Consider incorporating digital signatures or other authentication mechanisms to verify the integrity and authenticity of the encrypted data. Secure Transmission: If applicable, ensure secure transmission of encrypted files or images using protocols like HTTPS. Randomness: Ensure proper randomness in key generation and initialization vectors for AES to prevent vulnerabilities. Algorithm Updates: Stay informed about updates and best practices for the implemented algorithms to address potential vulnerabilities.

## **4 Results and Analysis**

Key Aggregate: Encryption Time and Decryption Time: Key Aggregate's encryption and decryption time could be more efficient than RSA and AES for small files due to its ability to handle multiple keys simultaneously. However, as file sizes increase, key aggregate might face challenges due to its design for aggregate key management, possibly causing encryption and decryption times to lag behind AES, which is optimized for larger data sizes. Encryption and

Decryption Throughput: Like time complexities, key aggregate might demonstrate higher throughput than RSA and AES for small files by efficiently handling multiple keys. Nevertheless, as file sizes grow, its throughput might diminish in comparison to AES, which is specifically designed for efficient processing of larger data chunks.

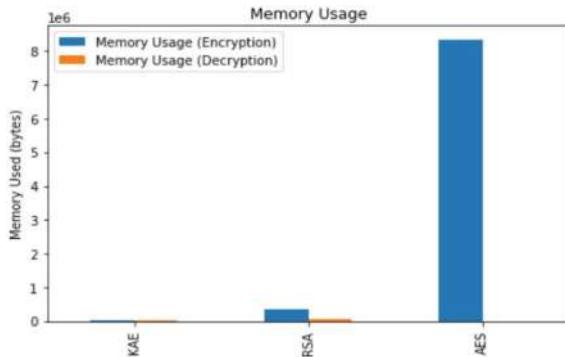
**Memory Used for Encryption and Decryption:** Key aggregate could potentially be more memory-efficient than RSA, particularly for small files, as it deals with aggregate key management, possibly reducing memory overhead. However, for larger files, its memory requirements might become less efficient compared to AES due to its design not being optimized for such scenarios. **Avalanche Effect for Small and Large Files:** For small files, key aggregate might not prioritize the avalanche effect as the primary concern, focusing instead on key aggregation efficiency. However, as file sizes increase, the avalanche effect could become less effective compared to AES, which maintains a robust avalanche effect even for larger data sets. **Time Complexity:** For key generation, time complexity is  $O(\log n)$ , and for encryption function and decryption function time complexity is  $O(n)$ . **Rivest–Shamir–Adleman (RSA): Time:** RSA involves complex mathematical operations like modular exponentiation, making it relatively slow compared to symmetric key algorithms like AES. The time complexity is higher, especially for key generation and encryption/decryption processes (Table 1).

**Throughput:** Due to its slower processing, RSA typically has lower throughput compared to symmetric key algorithms. **Memory Usage:** RSA uses larger key sizes for equivalent security, leading to increased memory usage. This can be a consideration in resource-constrained environments. **Avalanche Effect:** RSA exhibits a good avalanche effect, meaning that a small change in the input (either the plaintext or the key) results in a significantly different output. **Time complexity:** AES has  $O(n)$  as its time complexity. In summary, the choice between RSA, AES, and key aggregate depends on the specific requirements of the project. If speed and efficiency are paramount, AES is generally a better choice (Fig. 2).

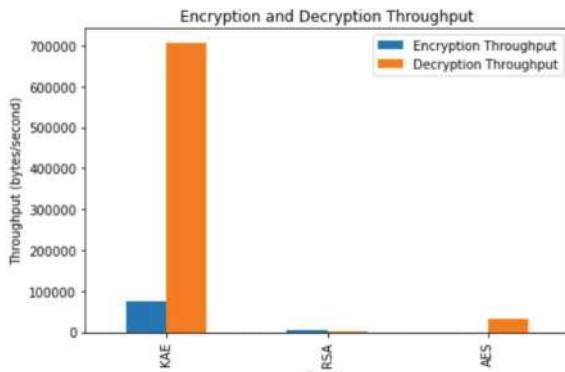
If key aggregation is a critical requirement, key aggregate may offer advantages in certain scenarios. RSA might be chosen for specific use cases where public-key cryptography is necessary despite its slower performance. The avalanche effect is generally strong in both AES and RSA, but its relevance might be lower in the context of key aggregation, where other factors might be more important. In summary, the key-aggregate algorithm shines in managing small files efficiently, surpassing both

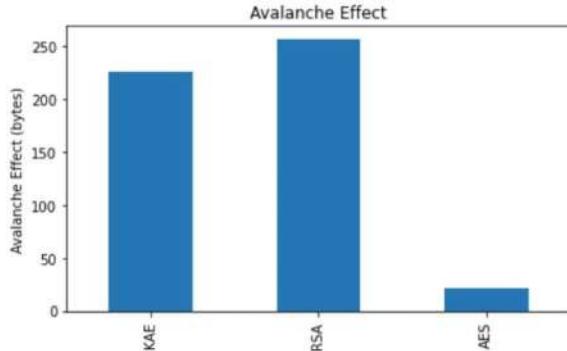
**Table 1** Encryption and decryption times

	KAE	RSA	AES
Encryption time	0.0002258 sec	0.0101418 sec	0.087713 sec
Decryption time	1.57e <sup>-05</sup> sec	0.0472206 sec	0.001188 sec
Input size	82 bytes	82 bytes	82 bytes
Memory used for decryption	18182144 bytes	344064 bytes	8335360 bytes
Memory used for decryption	15327194 bytes	49152 bytes	4096 bytes
Encryption throughout	200324.85 bytes/sec	3648.2497 bytes/sec	421.8256 bytes/sec
Decryption throughput	3775509.5 bytes/sec	772.7547 bytes/sec	31131.489 bytes/sec
Avalanche effect	2000 bytes	256 bytes	22 bytes

**Fig. 2** Memory usage

RSA and AES in certain aspects due to its specialized design for handling multiple keys. However, as file sizes increase, its advantages over RSA may persist, but its efficiency might start to diminish compared to AES. Additionally, the avalanche effect might become less effective in key-aggregate algorithm compared to RSA, especially with larger file sizes. Therefore, the choice of algorithm should consider the specific requirements of the file sizes and the desired balance between efficiency and cryptographic strength (Figs. 3 and 4).

**Fig. 3** Encryption and decryption throughput

**Fig. 4** Avalanche effect

## 5 Conclusion

Encryption is important in ensuring the security and privacy of digital communication, data, and information. The comparative analysis of RSA, AES, and key aggregate reveals that key aggregate exhibits superior performance in terms of time, throughput, and memory usage when compared to RSA and AES. The efficiency of key aggregate is particularly noteworthy in scenarios requiring key aggregation. However, it's essential to acknowledge that AES demonstrates a stronger avalanche effect, which enhances its cryptographic robustness compared to key aggregate, making AES more suitable for certain security-sensitive applications. Ultimately, the choice between these cryptographic algorithms should be driven by the specific requirements and priorities of the given project, considering factors such as efficiency, security, and the nature of the data being protected. In the context of comparing RSA, AES, and key aggregate for a project, a valuable future enhancement involves the integration of homomorphic encryption into the key aggregation process. Homomorphic encryption allows computations on encrypted data, enhancing security by enabling the aggregation of encrypted keys without disclosing their actual values. This approach ensures an additional layer of confidentiality during the aggregation process, preserving users' privacy and mitigating side-channel attacks. Furthermore, the implementation supports secure multi-party computation, facilitating joint computations on encrypted keys without revealing sensitive information. Evaluation metrics encompass security levels, performance overheads in terms of time, throughput, memory usage, resilience to attacks, and the practicality of implementing and maintaining homomorphic encryption in real-world scenarios. This enhancement offers a nuanced exploration of trade-offs between security and performance for a more comprehensive understanding of cryptographic approaches in the project.

## References

1. Cui B, Liu Z, Wang L (2021) Key-aggregate searchable encryption (KASE) for group data sharing via cloud storage. In: 2021 international conference on green energy, computing and sustainable technology (GECOST), Miri, Malaysia, pp 1–5. <https://doi.org/10.1109/TC.2015.2389959>
2. Liu J, Qin JA, Wang W (2023) Key-aggregate based access control encryption for flexible cloud data sharing. IEEE Access 11:20267–20283. <https://doi.org/10.1109/ACCESS.2023.3250106>
3. Susmitha C, Srinivasarao S, Laasya KS, Kannaiyah SK, Bulla S (2023) Hybrid cryptography for secure file storage. In: 2023 7th international conference on computing methodologies and communication (ICCMC), Erode, India, pp 1151–1156 <https://doi.org/10.1109/ICCMC56507.2023.10084073>
4. Jintcharadze E, Iavich M (2020) Hybrid implementation of twofish, AES, ElGamal, and RSA cryptosystems. In: 2020 IEEE east-west design test symposium (EWDTs), Varna, Bulgaria, pp 1–5. <https://doi.org/10.1109/EWDTs50664.2020.9224901>
5. Jasra B, Moon AH (2020) Image encryption techniques: a review. In: 2020 10th international conference on cloud computing, data science engineering (Confluence), Noida, India, pp 221–226, <https://doi.org/10.1109/Confluence47617.2020.9058071>
6. Yudheksa G, Kumar P, Keerthana S (2022) A study of AES and RSA algorithms based on GPUs, In: 2022 international conference on electronics and renewable systems (ICEARS), Tutticorin, India, pp 879–885. <https://doi.org/10.1109/ICEARS53579.2022.9752356>
7. Khanezaei N, Hanapi ZM (2014) A framework based on RSA and AES encryption algorithms for cloud computing services. In: 2014 IEEE conference on systems, process and control (ICSPC 2014), Kuala Lumpur, Malaysia, pp 58–62. <https://doi.org/10.1109/SPC.2014.7086230>
8. Soni D, Tiwari V, Kaur B, Kumar M (2021) Cloud computing security analysis based on RC6, AES and RSA algorithms in user-cloud environment. In: 2021 first international conference on advances in computing and future communication technologies (ICACFCT), Meerut, India, pp 269–273. <https://doi.org/10.1109/ICACFCT53978.2021.9837360>
9. Zhao N, Zhang G (2019) Privacy-protected certificateless aggregate signature scheme in VANET. In: 2019 11th international conference on wireless communications and signal processing (WCSP), Xi'an, China, pp 1–6. <https://doi.org/10.1109/WCSP.2019.8928040>
10. Chen W-H, Fan C-I, Tseng Y-F (2018) Efficient key-aggregate proxy reencryption for secure data sharing in clouds. In: 2018 IEEE conference on dependable and secure computing (DSC), Kaohsiung, Taiwan, pp 1–4. <https://doi.org/10.1109/DESEC.2018.8625149>
11. Yao Y, Zhai Z, Liu J, Li Z (2019) Lattice-based key-aggregate (Searchable) encryption in cloud storage. IEEE Access 7:164544–164555. <https://doi.org/10.1109/ACCESS.2019.2952163>
12. Liu Z, Liu Y (2018) Verifiable and authenticated searchable encryption scheme with aggregate key in cloud storage. In: 2018 14th international conference on computational intelligence and security (CIS), Hangzhou, China, pp 421–425. <https://doi.org/10.1109/CIS2018.2018.00100>
13. Jebaseeli KK, Rani VG (2019) Formulation of aggregate key in cloud by cipher text storage reduction for data outsourcing in healthcare. In: 2019 international conference on communication and electronics systems (ICCES), Coimbatore, India, pp. 828–833. <https://doi.org/10.1109/ICCES45898.2019.9002035>
14. Muthi Reddy P, Manjula SH, Venugopal KR (2018) Secured privacy data using multi key encryption in cloud storage. In: 2018 fifth international conference on emerging applications of information technology (EAIT), Kolkata, India, pp 1–4. <https://doi.org/10.1109/EAIT.2018.8470399>
15. Fan C-I, Tseng Y-F, Cheng-Yuan E, Huang J-J (2018) Transformation between attribute-based encryption and key-aggregate cryptosystem. In: 2018 13th Asia joint conference on information security (AsiaJCIS), Guilin, China, pp 35–41. <https://doi.org/10.1109/AsiaJCIS.2018.00015>
16. Kamimura M, Yanai N, Okamura S, Cruz JP (2020) Key-aggregate searchable encryption, revisited: formal foundations for cloud applications, and their implementation. IEEE Access 8:24153–24169. <https://doi.org/10.1109/ACCESS.2020.2967793>

17. Wang X, Cheng X, Xie Y (2020) Efficient verifiable key-aggregate keyword searchable encryption for data sharing in outsourcing storage. IEEE Access 8:11732–11742. <https://doi.org/10.1109/ACCESS.2019.2961169>
18. Solapurkar P (2015) Secure sharing of personal health records on cloud using key aggregate cryptosystem. In: 2015 international conference on information processing (ICIP), Pune, India, pp 278–283. <https://doi.org/10.1109/INFOP.2015.7489393>
19. Kendrekar PP, Chavhan MK (2016) Cryptographic implementation of aggregatekey encryption for data sharing in cloud storage. In: 2016 IEEE international conference on recent trends in electronics, information communication technology (RTEICT), Bangalore, India, pp 829–832. <https://doi.org/10.1109/RTE-ICT.2016.7807943>
20. Wang X., Xie Y., Cheng X., Jiang Z. (2019) An Efficient Key-Aggregate Keyword Searchable Encryption for Data Sharing in Cloud Storage. In: 2019 IEEE globecom workshops (GC Wkshps), Waikoloa, HI, USA, pp 1–6. <https://doi.org/10.1109/GCWkshps45667.2019.9024540>
21. Praveen I, Sethumadhavan M (2012) A more efficient and faster pairing computation with cryptographic security. In: Proceedings of the first international conference on security of internet of thing Aug 2012 <https://doi.org/10.1145/2490428.2490448>
22. Premalatha P, Premalatha P, Amritha PP, Amritha PP Optimally Locating for hiding information in audio signal. Int J Comput Appl (0975- 8887) 65(14):37–42
23. Menon RP Log Analysis Based Intrusion Prediction System. In: Emerging ICT for bridging the future proceedings of the 49th annual convention of the computer society of India (CSI)
24. Kumar AS, Girish KP A three factor authentication system for smart- card using biometric, visual cryptography and OTP. In: Artificial intelligence and evolutionary algorithms in engineering systems. Springer, India, pp 673–679

# Advancing Pneumonia Diagnosis: Hybrid and Optimal Deep CNN Model for Chest Image Classification



Gunapati Suresh , T. Ravi , and R. Krishnaprasanna

**Abstract** O Pneumonia is a significant worldwide health issue, with numerous etiologies, such as bacteria, fungi, and viruses, contributing to its challenging diagnosis. Computed tomography (CT) imaging is crucial for the diagnosis of pneumonia because it provides detailed information on lung abnormalities. During the pandemic, many hospitals have adopted CT scans to identify lung diseases caused by respiratory infections. Not only is it more expensive, but it's not as widely accessible. Early, precise, and cost-effective identification of pneumonia problems is necessary for improved treatment outcomes. Pneumonia-Plus, which uses convolutional neural network (CNN) architecture, was trained on a large set of annotated CT scans that included cases of bacterial, fungal, and viral pneumonia. We have optimized the system to independently recognize distinct features associated with multiple pneumonia causes. This work utilized a deep learning model, dubbed Pneumonia-Plus, to categorize viral, bacterial, and fungal pneumonia using chest photos. The experiment shows that Pneumonia-Plus performs better overall by classifying pneumonia cases with accuracy. Most importantly, the model achieves the best accuracy and recall rates while showing resilience in distinguishing between bacterial, viral, and fungal pneumonia.

**Keywords** C Pneumonia-plus · Chest image · Convolutional neural network · Deep learning

---

G. Suresh ()

Sathyabama Institute of Science and Technology, Chennai, India  
e-mail: [gsreddy455@gmail.com](mailto:gsreddy455@gmail.com)

T. Ravi · R. Krishnaprasanna

Department of Electronics and Communication Engineering, Sathyabama Institute of Science and Technology, Chennai, India

## 1 Introduction

Over the past few years, the medical field has encountered significant difficulties in addressing pneumonia, which is a major contributor to illness and death worldwide. The rise of new infectious pathogens, including SARS-CoV-2, in addition to the ongoing danger posed by microorganisms resistant to many drugs, highlights the urgent requirement for inventive methods of diagnosis and treatment. This study provides an in-depth analysis of the most recent progress in pneumonia care, including the discovery of underlying causes and the creation of advanced diagnostic tools and therapeutic approaches.

Pneumonia is a common and sometimes dangerous respiratory infection characterized by inflammation of the lungs, generally caused by bacteria, viruses, or fungi. It continues to be a prominent factor contributing to illness and death on a global scale, especially among susceptible groups such as children, the elderly, and people with weakened immune systems. An expedient and precise diagnosis of pneumonia is crucial for commencing suitable therapy and mitigating the risk of complications and mortality linked to the illness. Conventional approaches for diagnosing pneumonia involve assessing clinical symptoms, conducting physical examinations, and utilizing radiological imaging methods, such as chest X-rays (CXR) and CT images.

Nevertheless, effectively deciphering various imaging modalities might be a challenge, necessitating specialized training and experience. Moreover, the subjective nature of visually interpreting information might cause variations and inconsistencies in diagnosing conditions, which could potentially lead to misdiagnosis and delayed treatment. Due to recent progress in artificial intelligence (AI) and machine learning (ML) technologies, there is an increasing desire to create automated methods for diagnosing pneumonia using chest imaging. Deep learning, a subfield of machine learning, has become a potent tool for analyzing images, with the ability to acquire intricate patterns and characteristics straight from unprocessed data. Convolutional neural networks (CNNs) have demonstrated exceptional efficacy in several medical imaging applications, such as the identification of pneumonia from chest scans. Although deep learning techniques hold promise, there are still significant hurdles in creating sturdy and dependable CNN models for diagnosing pneumonia. An important obstacle is the scarcity of annotated medical image databases, especially for uncommon illnesses or specialized patient groups. Furthermore, the intrinsic complexity and diversity of chest images present difficulties in devising CNN architectures that can proficiently capture and interpret pertinent information for precise diagnosis. This paper presents a new hybrid CNN model as a solution to address these issues in diagnosing pneumonia from chest images.

The hybrid model combines various convolutional neural network architectures and utilizes optimization techniques to improve efficiency and the ability to generalize. The proposed model aims to address the limitations of individual CNN approaches and improve diagnostic accuracy by integrating the most effective components from several CNN designs and optimization techniques. In medical image processing, there are many deficiencies in the processing of CT images.

Pneumonia is a significant cause of illness and death globally, with several factors like bacteria, fungus, and viruses adding to its complexity. Conventional diagnostic techniques depend on clinical symptoms, physical examination, and radiological imaging. Computed tomography (CT) imaging provides intricate information about lung abnormalities, but its general use is limited due to its high cost and limited availability, especially in resource-limited settings. There is a requirement for accurate, economical, and early detection of pneumonia in order to improve treatment results and optimize the use of healthcare resources.

A literature review highlighted multiple research areas.

1. We must develop hybrid and optimal convolutional neural network (CNN) architectures capable of accurately distinguishing between bacterial, fungal, and viral pneumonia using chest pictures.
2. Utilize the Pneumonia-Plus model to train on a wide-ranging dataset of annotated CT scans that include several causes of pneumonia.
3. Enhance the functionality of Pneumonia-Plus to automatically extract unique characteristics linked to various types of pneumonia.
4. Assess the proficiency of Pneumonia-Plus in appropriately categorizing instances of pneumonia and differentiating between bacterial, fungal, and viral pneumonia.
5. The objective of the study is to evaluate the capacity of Pneumonia-Plus to accurately and promptly detect pneumonia problems in order to enhance treatment results.

## 2 Literature Review

The case study describes a 58-year-old man who had SARS-CoV-2 and developed hemophagocytic lymphocytosis (HLH). This case highlights the intricate nature of hyperserotonemia in severe sickness. Although the diagnosis of HLH was made, the extremely high levels of ferritin remained without explanation, emphasizing the significance of evaluating other possible diseases and carefully analyzing the increasing ferritin levels [1]. The prevalence of pneumonia remains a substantial health challenge, which is further aggravated by the increasing number of elderly individuals, the presence of chronic illnesses, and the emergence of new microbiological dangers. Successful management relies on comprehending the causes and patterns of microbial infections and resistance, emphasizing the significance of timely and suitable use of antimicrobial treatment in impacting the outcome. The task of matching broad-spectrum antibiotics with microbiological detection remains challenging, requiring a subtle and sophisticated approach to treatment [2]. The timely identification of pneumonia is crucial for better results, but the precise diagnosis remains difficult, especially in automated systems that do not have enough annotated data.

The novel Deep Supervised Domain Adaptation (DSDA) technique appears to be a promising approach for the detection of pneumonia through the analysis of chest X-rays. The performance is enhanced by utilizing extensive datasets and sub-networks

that are designed for certain tasks [3]. Deep learning algorithms have become significant instruments in the diagnosis of pneumonia, outperforming older methods in terms of accuracy and efficiency. Tests conducted with doctors as observers demonstrated that deep learning algorithms outperformed in classifying photos and detecting lesions, highlighting their significant utility for medical experts [4]. Fungal pneumonia in elderly people is frequently associated with elevated mortality rates and presents diagnostic difficulties. This case presentation underscores the need of considering fungal pneumonia in instances that do not respond to treatment, underlining the necessity of using thorough diagnostic methods that go beyond the use of broad-spectrum antibiotics [5].

Chest X-ray pictures are analyzed using advanced convolutional neural network (CNN) models to directly extract characteristics and detect pneumonia. This approach effectively tackles concerns related to the reliability and interpretability of the results. Data augmentation methods improve the validation and accuracy of models, representing the possible of AI in the field of medical imaging [6]. Sarcopenia is a vital imaging biomarker that can envisage clinical results. Different modalities can afford data on muscle mass and superiority. Radiologists have a vital role in diagnosing sarcopenia, which requires them to have knowledge and conduct research on the best imaging techniques and thresholds [7]. During the COVID-19 pandemic, it is crucial to have efficient and accurate diagnostic tools. Deep learning frameworks, which are based on advanced algorithms, provide precise and effective detection of COVID-19 from chest radiographs. This complements the current diagnostic methods and helps in allocating resources efficiently [8]. Despite the inherent limitations in image quality, automated systems explicitly designed for classifying chest X-ray images into different categories like normal, abnormal, and COVID-19 categories show promising accuracy. Deep learning methods improve the ability to distinguish and aid in making therapeutic decisions [9]. An analysis of serum samples from COVID-19 patients reveals alterations in molecular composition associated with severe illness, indicating that blood biomarkers may serve as indicators of disease severity and aid in early detection [10]. The examination of chest CT results in individuals with COVID-19 uncovers frequent anomalies and their association with clinical characteristics, highlighting the crucial significance of CT in diagnosing and assessing the condition [11]. Automated deep learning systems, which analyze regularly performed CT scans, effectively detect COVID-19 cases and categorize patients according to their risk level. This helps in allocating resources and intervening early [12]. Chest CT imaging is essential for diagnosing COVID-19, as it helps distinguish it from other types of pneumonia using specific imaging patterns [13]. CADx systems provide precise classification of COVID-19 pneumonia from chest X-ray images, offering important assistance to radiologists in diagnosis and research [14]. AI models surpass radiologists in differentiating COVID-19 from other kinds of pneumonia on chest CT scans, emphasizing the potential of AI in enhancing diagnostic precision [15]. Attention-based convolutional neural networks improve the accuracy of diagnosing COVID-19 from chest X-ray pictures by overcoming the limitations of current methods [16]. Deep learning models that can do many tasks simultaneously are used to enhance the accuracy of COVID-19 pneumonia screening

by chest CT imaging. These models improve the segmentation, classification, and reconstruction tasks, resulting in improved accuracy [17].

Deep learning algorithms exhibit a notable level of specificity in detecting COVID-19-related pneumonia and differentiating it from other types of pneumonia. This assists in achieving precise diagnoses in various groups of patients [18]. By incorporating both anomaly and confidence scores, confidence-aware anomaly detection models demonstrate enhanced capability in identifying instances of viral pneumonia [19]. Wavelet decomposition in enhanced depth-wise convolution neural networks has been shown to enhance diagnostic accuracy for COVID-19, making it a potential approach for illness diagnosis [20]. Combining information from frontal and lateral X-ray pictures in a neural network improves the accuracy of diagnosing pneumothorax, making it a potential approach for medical imaging [21]. Transfer learning methodologies are used to enhance the accuracy and efficiency of COVID-19 detection from X-rays using deep learning methods [22]. A novel framework for segmenting COVID-19 CXR images improves the representation of features and semantic linkages, resulting in better performance compared to current networks [23]. Explainable AI (XAI) allows for the clear and understandable categorization of pneumonia, providing flexible and precise diagnostic capacities that are crucial for the implementation of healthcare systems worldwide [24]. Advanced imaging techniques such as 3D X-ray Scatter Tomography facilitate the visualization of lung structures, potentially revolutionizing the diagnostic process without increasing radiation exposure [25]. The SC2Net, a network that uses segmentation to classify COVID-19, shows exceptional performance in diagnosing COVID-19 from chest X-ray images. It effectively tackles the issues of early-stage diagnosis [26]. Deep neural networks present a highly promising method for automating the diagnosis of COVID-19, simplifying interpretation and assisting in the classification of the condition [27].

Cycle Generative Adversarial Networks (GANs) facilitate the visualization of the progression of pneumonia by redefining classifiers and generating realistic images that depict the worsening of the disease. This aids in improving the accuracy of diagnosis [28]. Deep learning diagnostic technologies provide clear and precise diagnosis, as well as simplify the process of referring patients, resulting in improved clinical outcomes for curable illnesses like pneumonia [29]. The utilization of chest CT images in conjunction with the Pneumonia-Plus deep learning algorithm enables the precise categorization of bacterial, fungal, and viral pneumonia. This categorization aids doctors in determining the appropriate treatment for patients and reduces the likelihood of incorrect diagnoses [30].

The combination of modern imaging techniques, artificial intelligence, and deep learning algorithms has the potential to greatly transform the diagnosis and treatment of pneumonia. These novel strategies provide healthcare professionals with essential resources to effectively handle the intricacies of pneumonia treatment, thereby enhancing patient outcomes and reducing the worldwide impact of this widespread infectious illness. A literature review highlighted several research areas.

We must develop hybrid and optimal convolutional neural network (CNN) architectures capable of utilizing chest pictures to accurately distinguish between bacterial, fungal, and viral pneumonia.

Utilize the Pneumonia-Plus model to train on a wide-ranging dataset of CT scans that have been annotated and cover various causes of pneumonia. Enhance the functionality of Pneumonia-Plus to automatically extract unique characteristics linked to various types of pneumonia. Assess the efficacy of Pneumonia-Plus in accurately categorizing cases of pneumonia and differentiating between bacterial, fungal, and viral pneumonia. The objective of the study is to evaluate the capacity of Pneumonia-Plus to accurately and promptly detect pneumonia problems in order to enhance treatment results.

### 3 Related Works

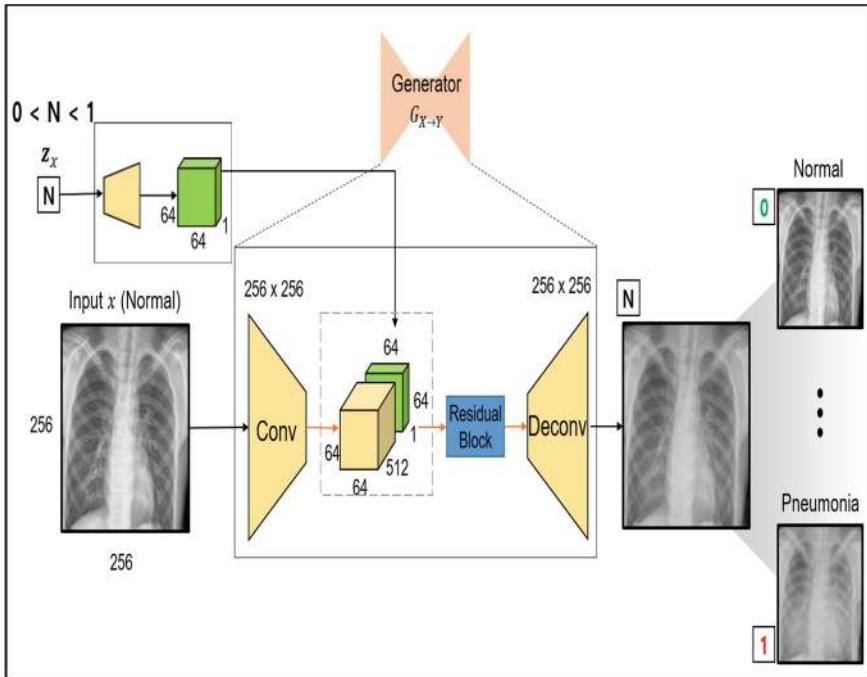
This section uses established procedures to conduct a thorough analysis.

#### 3.1 Using Conditional CycleGANs

Scientists created conditional CycleGANs [27] to imitate the progression of pneumonia through the production of images. Conditional CycleGANs enhance the conventional CycleGAN framework by including conditional vectors that guide the picture alteration process across different domains. This improvement enhances the level of precision in controlling the generated images according to specified conditions.

This architectural design incorporates conditional vectors within the mapping function, guaranteeing both cycle consistency and the preservation of identity, all based on the given conditional vector. Cycle consistency refers to the property that a picture, when transformed from one domain to another and then transformed back to the original domain, should maintain its original attributes. After undergoing processing through the generators in a cycle, an image  $x$  from the normal domain  $X$  should remain unchanged and identical to the original image. Figure 1 depicts the process wherein the generators  $G_X$  and  $G_Y$  are used to translate an image  $x$  from domain  $X$  to domain  $Y$  and then back to domain  $X$ . The objective is to ensure that the resulting output image is identical to the original  $x$ , making it impossible to discriminate between them.

The mathematical representation of this is  $G_Y \rightarrow X (G_X \rightarrow Y (x, c_1), c_0) = x$ , where  $c_0$  and  $c_1$  are the conditional inputs for normal and pneumonia pictures, respectively. This equation guarantees that the resulting image, after undergoing two consecutive translations between the domains, preserves the faithfulness to the original image, with the conditional inputs  $c_0$  and  $c_1$  directing the transformations.



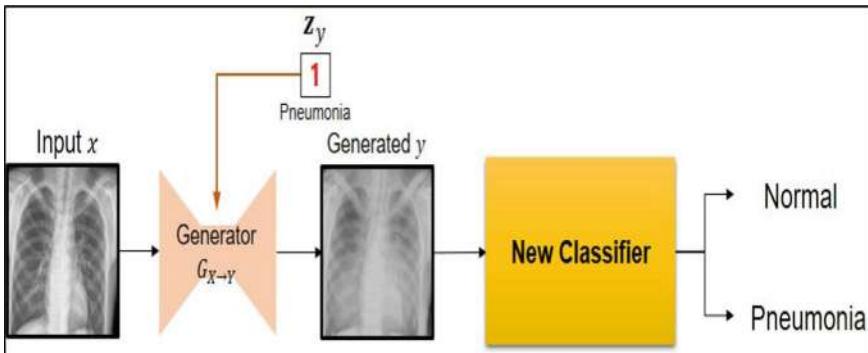
**Fig. 1** Generator structure for generating pneumonia progression

In order to accomplish this, researchers utilize a conditional cycle consistency loss, which guarantees that images that undergo these modifications maintain their original form when cycled back. The loss function is essential for preserving the structural and visual integrity of the images, which makes the conditional CycleGAN a highly effective tool for predicting illness development. Figure 1 depicts the generator architecture used to advance the technique [27].

This methodology replicates the development of pneumonia in medical imaging, providing a vital tool for medical research and diagnostics. By utilizing conditional vectors, one may effectively manipulate picture attributes, enabling a meticulous examination of the progression of pneumonia over time. This breakthrough not only enhances the comprehension of the condition, but it also expands the potential for creating diagnostic tools and treatment strategies based on a visual examination of disease progression. Figures 2 and 3 depict the classification for regular test images and test images that have undergone a shift in domain, respectively [27].



**Fig. 2** Classification for test images [27]



**Fig. 3** Classification for domain changed test images [27]

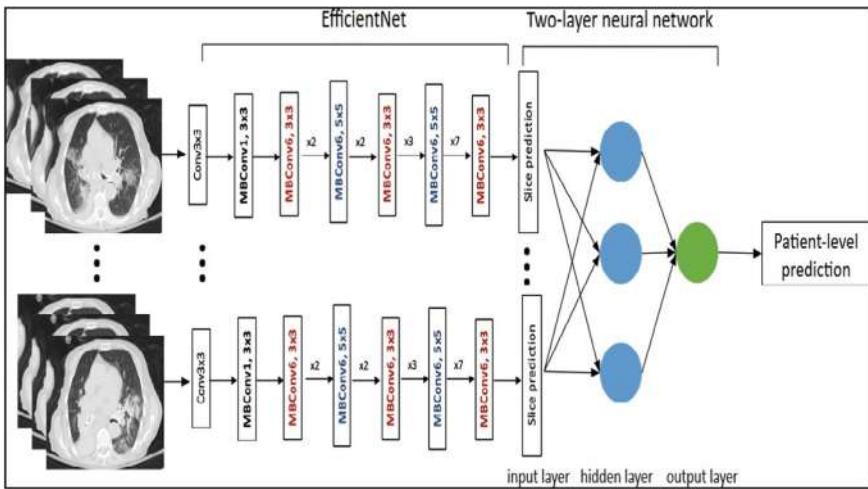
### 3.2 Using CT Imaging for COVID-19

Bai et al. [14] provide a thorough explanation of the pneumonia classification model utilizing the EfficientNet architecture, as demonstrated in the CT images below.

The authors trained a classification model to distinguish between CT scan slices showing pneumonia-like features and those that do not. This included both COVID-19 and non-COVID-19 cases. The selected the EfficientNet design due to its use of mobile inverted bottleneck MBConv blocks. These blocks decrease model parameters and enhance accuracy and efficiency over traditional convolutional networks. The EfficientNet-B3 model, pre-trained on ImageNet, featured a single fully connected layer for two-class classification. This layer had a dropout probability of 0.5.

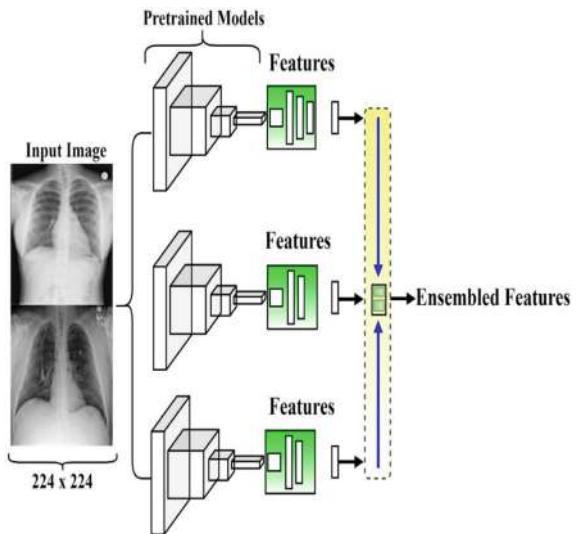
During training, the authors implemented dynamic data augmentation using various transformations. These included flips, scaling, rotations, brightness and contrast modifications, random noise, and blurring. The training process lasted for 20 epochs, with each epoch consisting of 16,000 slices. They used the AdamW optimizer with its default settings. They performed bi-epoch validation and assessed performance by measuring the area under the curve (AUC).

Figure 4 shows the COVID-19 Classification Neural Network Model. Figure 5 illustrates the flow diagram of the AI model used to differentiate between COVID-19 and non-COVID-19 pneumonia. The ultimate model selected was the checkpoint that had the largest validation AUC (Area Under the Curve). The decisions



**Fig. 4** COVID-19 classification neural network model [14]

**Fig. 5** Diagrammatical illustration of the proposed ensemble architecture [31]



on network scaling were made based on empirical evidence of the validation set's performance, taking into account the trade-off between size and performance, as well as computational efficiency.

The Efficient Net B4 architecture was employed for the job of classifying pneumonia. The input slices were arranged in three channels in order to utilize the pre-trained ImageNet weights effectively. The Efficient Net models were enhanced by adding dense top fully connected layers consisting of four layers of 256, 128, 64, and 32 neurons, respectively. These layers were equipped with a dropout rate of

0.5, ReLU activations, and batch normalization. The neural network consisted of a fully connected layer including 16 neurons, which was then followed by a classification layer containing a single neuron activated by the sigmoid function. They used this network to predict whether the given slices were related to COVID-19 or non-COVID-19 pneumonia. A two-layer fully connected neural network made patient-level predictions by pooling the slice predictions. The authors used stochastic gradient descent as the optimizer, with a learning rate of 0.0001 and a batch size of 64. Grad-CAM was used to create heatmaps of significant image regions that influenced the model's categorization. Thus, the compared demographic and clinical features between COVID-19 pneumonia and non-COVID-19 pneumonia groups using chi-square tests for categorical variables and Student t-tests for continuous variables. The Mann–Whitney U test compared CT slice thickness between the COVID-19 and non-COVID-19 groups. Additionally, the Kruskal–Wallis H test compared slice thickness among the training, validation, and test sets [14].

To evaluate the performance of the classification model, we used the adjusted Wald technique to calculate the 95% confidence intervals for accuracy, sensitivity, and specificity. The authors conducted a benchmark of the model's performance in comparison with the average performance of radiologists. Additionally, we employed permutation tests to assess the performance of radiologists with and without the assistance of AI. The performed studies utilizing the *R* statistical computing language are used for future researchers [14].

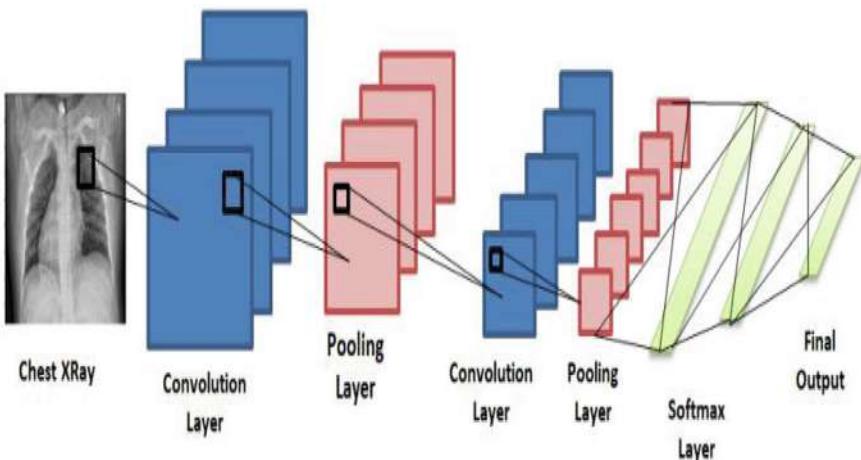
Recent methodologies have uncovered a promising field of study concerning CT imaging of pneumonia and related lung illnesses. Preprocessing, datasets, optimization methodologies, evaluation, and validation are all essential prerequisites for the next research.

### ***3.3 Multi-model Ensemble Deep Learning***

The reference [31] clearly explained about Multi-model ensemble deep learning for Chest X-Ray images. The architecture is as in Fig. 6 (Table 1).

### ***3.4 Efficient Pneumonia Detection in Chest X-Ray Images Using Deep Transfer Learning [32]***

In the reference [32], a technique was proposed weighted classifier demonstrates superior performance compared to each of the individual models. Comprehensive evaluation of the model is conducted, assessing not only its test accuracy but also its AUC (Area Under the Curve) score. The final weighted classifier achieves a test accuracy of 98.43% and an AUC score of 99.76% on unseen data from the Guangzhou Women and Children's Medical Center pneumonia dataset. These results suggest that



**Fig. 6** Architecture of pneumonia detection in chest X-ray images using deep transfer learning

**Table 1** Chest X-ray dataset distribution over binary and multiclass classification scenarios

Dataset	Partition	Normal	Viral pneumonia	Bacteria pneumonia	Total/partition	Total
Mendeley: binary task: two classes	Train	3600	3900	–	7500	8124
	Test	234	390	–	624	
Chest X-ray: multiclass task: three classes	Train	3500	3500	3500	10,500	15,000
	Test	1500	1500	1500	4500	

Training and validation images are combined for fine-tuning the deep learning models

the proposed model is highly effective for the rapid diagnosis of pneumonia and can serve as a valuable tool to assist radiologists in the diagnostic process. The results of the above architecture are tabulated in Tables 2 and 3.

**Table 2** Final testing accuracy and testing loss achieved by all the architectures and the weighted classifier

Architecture	Testing accuracy	Testing loss
ResNet18	97.29	0.096
DenseNet121	98.00	0.064
Inception	97.00	0.098
Xception	96.57	0.101
MobileNetV2	96.71	0.096
Weighted classifier (with equal weights)	97.45	0.087
Weighted classifier (with optimized weights)	98.43	0.062

**Table 3** Weight value (belief or trust value) corresponding to every architecture

Architecture	Weight
ResNet18 (W1)	0.25
DenseNet121 (W2)	0.30
Inception (W3)	0.18
Xception (W4)	0.08
MobileNetV2 (W5)	0.19

## 4 Essential Steps for Methodology

The essential steps to include for an efficient are explained in this section.

### 4.1 Data Collection and Preprocessing

Collect a comprehensive dataset of annotated CT scans encompassing bacterial, fungal, and viral pneumonia cases from diverse sources. Preprocessing the CT scan images to standardize resolution, normalize intensities, and remove noise and artifacts. To increase diversity and improve model generalization, augment the dataset with techniques such as rotation, flipping, and scaling.

### 4.2 Architecture Design

Develop an efficient and optimum convolutional neural network (CNN) structure for Pneumonia-Plus by integrating conventional CNN layers with state-of-the-art feature extraction methods. Conduct experiments with various architectures, such as residual

connections, inception modules, and attention algorithms, to improve the representation of features and the performance of categorization. Apply transfer learning by leveraging pre-trained models learned on extensive image datasets. Figure 1 depicts the flow chart of the proposed approach.

### ***4.3 Training and Optimization***

Utilize stochastic gradient descent (SGD) or adaptive optimization methods to train Pneumonia-Plus on the annotated CT scan dataset and determine the optimal model parameters. Utilize strategies such as learning rate scheduling, batch normalization, and dropout regularization. In order to enhance the performance of the model, it is advisable to adjust the hyperparameters of the model based on the validation performance. This includes optimizing the learning rate, batch size, and dropout rate.

### ***4.4 Evaluation and Validation***

Evaluate the performance of Pneumonia-Plus on distinct test sets by analyzing metrics like as F1-score, accuracy, recall, precision, and area under AUC-ROC, the receiver operating characteristic curve. Assess the resilience of Pneumonia-Plus in different pneumonia causes and patient characteristics, guaranteeing its applicability to real-life clinical situations. To showcase its advantage in diagnosing pneumonia, we can compare the performance of Pneumonia-Plus with both existing deep learning models and traditional diagnostic approaches.

### ***4.5 Clinical Application and Impact Assessment***

Implement the Pneumonia-Plus tool in clinical environments and incorporate it into current healthcare systems to facilitate the diagnosis of pneumonia. Perform clinical trials and case studies to evaluate the effects of Pneumonia-Plus on treatment results, utilization of healthcare resources, and patient care. Evaluate the practical value of Pneumonia-Plus by implementing it in real-world settings and gathering input from healthcare professionals and patients.

## 4.6 Ethical Considerations

Guarantee adherence to ethical rules for the collecting of data, creation of models, and implementation in clinical settings. Ensure the security of patient privacy and confidentiality during the research process, by strictly following applicable data protection legislation. Ensure that patients who are involved in clinical trials or contributing medical data for research reasons provide their informed permission.

## 5 Conclusion

Pneumonia-Plus is an innovative development in the diagnosis of pneumonia, providing a precise, cost-effective, and early detection tool for various causes of pneumonia. Pneumonia-Plus employs a hybrid and optimal convolutional neural network (CNN) structure, together with a substantial dataset of annotated CT scans, to enhance its ability to accurately categorize pneumonia cases and distinguish between bacterial, fungal, and viral pneumonia. The model's resilience and exceptional accuracy and sensitivity rates highlight its potential as a significant tool for doctors in diagnosing pneumonia and devising treatment strategies.

Moreover, the effective implementation of Pneumonia-Plus in clinical environments has the capacity to simplify healthcare processes, enhance treatment results, and maximize the use of healthcare resources. Pneumonia-Plus improves patient care and supports the global fight against pneumonia-related illness and death by giving doctors prompt and precise pneumonia diagnoses.

## References

1. Cilloniz C, Martin-Loches I, Garcia-Vidal C, San Jose A, Torres A (2016) Microbial etiology of pneumonia: epidemiology, diagnosis and resistance patterns. *Int J Mol Sci* 17(12):2120. <https://doi.org/10.3390/ijms17122120>
2. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F, Dong J, Prasadha MK, Pei J, Ting MYL, Zhu J, Li C, Hewett S, Dong J, Ziyar I, Shi A, Zhang R, Zheng L, Hou R, Shi W, Fu X, Duan Y, Huu VAN, Wen C, Zhang ED, Zhang CL, Li O, Wang X, Singer MA, Sun X, Xu J, Tafreshi A, Lewis MA, Xia H, Zhang K (2018) Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* 172(5):1122-1131.e9. <https://doi.org/10.1016/j.cell.2018.02.010>
3. Stephen O, Sain M, Maduh UJ, Jeong DU (2019) An efficient deep learning approach to pneumonia classification in healthcare. *J Healthc Eng* 2019:4180949. <https://doi.org/10.1155/2019/4180949>
4. Hwang EJ, Park S, Jin KN, Kim JI, Choi SY, Lee JH, Goo JM, Aum J, Yim JJ, Cohen JG, Ferretti GR, Park CM, DLAD Development and Evaluation Group (2019) Development and validation of a deep learning-based automated detection algorithm for major thoracic diseases on chest radiographs. *JAMA Netw Open* 2(3):e191095 (2019). <https://doi.org/10.1001/jamaneurology.2019.1095>

5. Yang W, Ma Y, Wang J, Li Y, Zhang Y, Zhang T (2019) Fungal pneumonia manifesting as cavitary lesions in a critically Ill elderly patient. *J Infect Dev Ctries* 13(12):1170–1173. <https://doi.org/10.3855/jidc.11265>
6. Li X, Zeng W, Li X, Chen H, Shi L, Li X, Xiang H, Cao Y, Chen H, Liu C, Wang J (2020) CT imaging changes of corona virus disease 2019 (COVID-19): a multi-center study in Southwest China. *J Transl Med* 18(1):154. <https://doi.org/10.1186/s12967-020-02324-w>
7. Albano D, Messina C, Vitale J et al (2020) Imaging of sarcopenia: old evidence and new insights. *Eur Radiol* 30:2199–2208. <https://doi.org/10.1007/s00330-019-06573-2>
8. Sakib S, Tazrin T, Fouad MM, Fadlullah ZM, Guizani M (2020) DL-CRC: deep learning-based chest radiograph classification for COVID-19 detection: a novel approach. *IEEE Access* 8:171575–171589. <https://doi.org/10.1109/ACCESS.2020.3025010>
9. De Moura J et al (2020) Deep convolutional approaches for the analysis of COVID-19 using chest X-ray images from portable devices. *IEEE Access*. 8:195594–195607. <https://doi.org/10.1109/ACCESS.2020.3033762>
10. Arias-Londono JD, Gomez-Garcia JA, Moro-Velazquez L, Godino-Llorente JI (2020) Artificial intelligence applied to chest X-ray images for the automatic detection of COVID-19. A thoughtful evaluation approach. *IEEE Access*. 8:226811–226827. <https://doi.org/10.1109/ACCESS.2020.3044858>
11. Shen B, Yi X, Sun Y, Bi X, Du J, Zhang C, Quan S, Zhang F, Sun R, Qian L, Ge W, Liu W, Liang S, Chen H, Zhang Y, Li J, Xu J, He Z, Chen B, Wang J, Yan H, Zheng Y, Wang D, Zhu J, Kong Z, Kang Z, Liang X, Ding X, Ruan G, Xiang N, Cai X, Gao H, Li L, Li S, Xiao Q, Lu T, Zhu Y, Liu H, Chen H, Guo T (2020) Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell* 182(1):59–72.e15. <https://doi.org/10.1016/j.cell.2020.05.032>
12. Wu J, Wu X, Zeng W, Guo D, Fang Z, Chen L, Huang H, Li C (2020) Chest CT findings in patients with coronavirus disease 2019 and its relationship with clinical features. *Invest Radiol* 55(5):257–261. <https://doi.org/10.1097/RLI.0000000000000670>
13. Nishio M, Noguchi S, Matsus H et al (2020) Automatic classification between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy on chest X-ray image: combination of data augmentation methods. *Sci Rep* 10:17532. <https://doi.org/10.1038/s41598-020-74539-2>
14. Bai HX, Wang R, Xiong Z, Hsieh B, Chang K, Halsey K, Tran TML, Choi JW, Wang DC, Shi LB, Mei J, Jiang XL, Pan I, Zeng QH, Hu PF, Li YH, Fu FX, Huang RY, Sebro R, Yu QZ, Atalay MK, Liao WH (2020) Artificial intelligence augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other origin at chest CT. *Radiology* 296(3):E156–E165. <https://doi.org/10.1148/radiol.2020201491>
15. Harmon SA, Sanford TH, Xu S et al (2020) Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. *Nat Commun* 11:4080. <https://doi.org/10.1038/s41467-020-17971-2>
16. Wang S, Zha Y, Li W, Wu Q, Li X, Niu M, Wang M, Qiu X, Li H, Yu H, Gong W, Bai Y, Li L, Zhu Y, Wang L, Tian J (2020) A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *Eur Respir J* 56(2):2000775. <https://doi.org/10.1183/13993003.00775-2020>
17. Amyar A, Modzelewski R, Li H, Ruan S (2020) Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: classification and segmentation. *Comput Biol Med* 126:104037. <https://doi.org/10.1016/j.combiomed.2020.104037>
18. Zellweger NM, Huber J, Tsakiris DA, Tzankov A, Gebhard CE, Siegemund M (2021) Haemophagocytic Lymphohistiocytosis and liver failure-induced massive Hyperferritininaemia in a male COVID-19 patient. *Swiss Med Wkly* 151:w20420. <https://doi.org/10.4414/smw.2021.20420>
19. Zhang J, Xie Y, Pang G, Liao Z, Verjans JW, Li W, Sun Z, Shen C, Xia Y (2021) Viral pneumonia screening on chest X-rays using confidence-aware anomaly detection. *IEEE Trans Med Imaging* 40(3):879–890. <https://doi.org/10.1109/TMI.2020.3040950>
20. Rundo L, Militello C, Russo G, Vitabile S, Gilardi MC, Mauri G (2021) Automated prostate gland segmentation based on histograms of object-oriented gradients and probabilistic voting strategy. *Int J Comput Assist Radiol Surg* 16:645–655. <https://doi.org/10.1007/s11548-021-02357-w>

21. Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, Ahmadi M, Kalhori SRN, Hosseini NS (2020) Predicting COVID-19 incidence through analysis of google trends data in Iran: data mining and deep learning pilot study. *JMIR Public Health Surveill* 6(2):e18828. <https://doi.org/10.2196/18828>
22. Signoroni A, Savardi M, Benini S, Adami N, Leonardi R, Gibellini P, Vaccher F, Ravanelli M, Borghesi A, Maroldi R (2020) End-to-end learning for semantics-guided lung nodule detection in chest radiograms. *IEEE Trans Med Imaging* 39(8):2924–2935. <https://doi.org/10.1109/TMI.2020.2979561>
23. Feng Y, Yang X, Qiu D, Zhang H, Wei D, Liu J (2022) PCXRNet: pneumonia diagnosis from chest X-ray images using condense attention block and multiconvolution attention block. *IEEE J Biomed Health Inform* 26(4):1484–1495. <https://doi.org/10.1109/JBHI.2022.3148317>
24. Feng Y, Yang X, Qiu D, Zhang H, Wei D, Liu J (2022) Deep supervised domain adaptation for pneumonia diagnosis from chest X-ray images. *IEEE J Biomed Health Inform* 26(3):1080–1090. <https://doi.org/10.1109/JBHI.2021.3100119>
25. Ruben G, Pinar I, Brown JMC, Schaff F, Pollock JA, Crossley KJ, Maksimenko A, Hall C, Hausermann D, Uesugi K, Kitchen MJ (2022) Full field X-ray scatter tomography. *IEEE Trans Med Imaging* 41(8):2170–2179. <https://doi.org/10.1109/TMI.2022.3157954>
26. Zhao H, Xu C, Wei Z, Yan B, Huang Y, Zhang J, Ding J, Zhao J, Zhang H (2022) SC2Net: a novel segmentation-based classification network for detection of COVID-19 in chest X-ray images. *IEEE J Biomed Health Inform* 26(8):4032–4043. <https://doi.org/10.1109/JBHI.2022.3177854>
27. Jin Y, Chang W, Ko B (2023) Generating chest X-ray progression of pneumonia using conditional cycle generative adversarial networks. *IEEE Access* 11:88152–88160. <https://doi.org/10.1109/ACCESS.2023.330599>
28. Sheu R-K, Pardeshi MS, Pai K-C, Chen L-C, Wu C-L, Chen W-C (2023) Interpretable classification of pneumonia infection using eXplainable AI (XAI-ICP). *IEEE Access* 11:28896–28919. <https://doi.org/10.1109/ACCESS.2023.3255403>
29. Yue G, Lin J, An Z, Yang Y (2023) Loop residual attention network for automatic segmentation of COVID-19 chest X-ray images. *IEEE Access* 11:47480–47490. <https://doi.org/10.1109/ACCESS.2022.3227798>
30. De Castro Santos MA, Berton L (2023) An enhanced framework for overcoming pitfalls and enabling model interpretation in pneumonia and Covid-19 classification. *IEEE Access* 11:115330–115347. <https://doi.org/10.1109/ACCESS.2023.3325404>
31. Chiagoziem C et al (2023) A hybrid explainable ensemble transformer encoder for pneumonia identification from chest X-ray images. *J Adv Res* 48:191–211. <https://doi.org/10.1016/j.jare.2022.08.021>
32. Hashmi MF, Katiyar S, Keskar AG, Bokde ND, Geem ZW (2020) Efficient pneumonia detection in chest X-ray images using deep transfer learning. *Diagnostics* 10(6):417. <https://doi.org/10.3390/diagnostics10060417>

# Enhancing Intrusion Detection by Integrating Deep Learning and Traditional Machine Learning Techniques



Noor Yeshfeen and Ritika Kumari

**Abstract** The main goal of this project is to solve the problem of network intrusion detection system, we have used deep learning and traditional machine learning techniques to enhance intrusion detection. Intrusion detection plays a very important role in network security, which basically helps to detect malicious activities or anomaly behavior by monitoring network traffic. Or it has used traditional techniques but due to its limitations it has become necessary to integrate deep learning. In the project we used the NSL-KDD dataset, which is a widely accepted dataset for network traffic and intrusion data. Feature extraction and dimensionality reduction techniques such as PCA (Principal Component Analysis) have been used to process the data. In the algorithm we used decision trees, random forest, SVM, ANN and ensemble methods, which improved our detection accuracy. The future scope of this project is to use hybrid approach in real-time network security systems, so as to reduce false alarms along with high detection accuracy. Going forward, we can use it in distributed systems and integrate advanced deep learning models, so that even large and complex networks can be effectively secured.

**Keywords** Network intrusion detection · Traditional ML · DL techniques · Feature scaling · Model evaluation · Principal component analysis (PCA) · Dimensionality reduction · Performance metrics · Accuracy · Precision · Recall · Loss computational efficiency · Intrusion detection systems

---

N. Yeshfeen (✉) · R. Kumari

Department of Artificial Intelligence and Data Sciences, IGDTUW, New Delhi, Delhi, India  
e-mail: [Noor023mtaids22@igdtuw.ac.in](mailto:Noor023mtaids22@igdtuw.ac.in)

R. Kumari  
e-mail: [ritikakumari@igdtuw.ac.in](mailto:ritikakumari@igdtuw.ac.in)

## 1 Introduction

Our project is based on “Enhancing Intrusion Detection by Integrating Deep Learning and Traditional Machine Learning Techniques”. Or the system is being used a lot in today’s time or network attacks like anomaly attacks are increasing in today’s digital world. As per our project, we have used our intrusion detection to detect malicious activities or attacks. Moreover, in this project, we have used machine learning methods and multiple models or have created a hybrid model by combining deep learning techniques. The traditional methods are fast and explainable, but do not detect complex patterns. But deep learning algorithms can effectively learn complex patterns which are necessary to detect modern sophisticated attacks. Or in the near future, we can implement our hybrid model on real-time network traffic, which can lead to the development of automatic intrusion detection systems, all of which can further strengthen network security.

### Motivation

The main motivation of our project is to improve network security, in which intrusion detection has been used to improve intrusion detection. With the integration of machine learning techniques and deep learning approaches, our main aim is to be able to detect malicious activities in real-time in a very accurate manner. Considering the sophisticated cyber-attacks of today, it is necessary to create a system which can accurately detect both known and unknown attacks or threats. The motivation of this project is to enhance the performance of intrusion detection so that safe communication can be ensured in network environments.

### Main Contribution

The main contribution is:

**Hybrid Approach for Intrusion Detection:** In this project we have integrated traditional machine learning algorithms along with the deep learning techniques. This hybrid approach helps in increasing the accuracy and efficiency of intrusion detection.

**Data Imbalance Handling:** We have implemented specific techniques to handle unbalanced datasets, so that even rare attack patterns can be efficiently detected.

**Feature Extraction and Dimensionality Reduction:** Here we have used PCA techniques so that high-dimensional data can be made low dimensional or even simplified, without losing any important information. With the help of PCA techniques the computational efficiency of the system is increased.

**Real-Time Detection Capability:** Our system is designed to perform real-time intrusion detection, allowing network security systems to receive real-time alerts and responses.

**Performance Optimization:** We have trained and tested different models to improve performance metrics like accuracy, precision, recall, F1-score and AUC, so that an optimal solution can be developed.

## Organization of the Paper

The organization of the paper are as follows:

**Section 2: Related Work:** This segment examines the current body of research on intrusion detection systems, highlighting the strengths and limitations of traditional and contemporary approaches.

**Section 3: Research Methodology:** We outline our suggested approach, which encompasses data preparation, feature selection, and the combination of deep learning with conventional machine learning methods.

**Section 4: Results and Discussion:** This describes the experimental setup, datasets used, and the results of our comprehensive evaluations, including performance metrics and comparisons with existing methods.

**Section 5: Conclusion:** We explore the significance of our results, possible constraints, and avenues for future investigation.

### 1.1 Problem Statement

The main goal in this project is to achieve high accuracy while maintaining efficiency. Traditional ml methods, like decision trees and SVM, are efficient but may not capture complex patterns. On the other hand, deep learning methods can capture intricate patterns but are computationally intensive.

## 2 Related Work

Different interruption discovery frameworks have utilized the Data Discovery 99 and Network Security Logs Dataset to assess their execution plus adequacy in the writing. For occasion, Ingre et al. [1] formulated a 3-Stacked Neural Network distinguish assault categories Network Security Logs Dataset inside the NSL-KDD dataset, accomplishing a multi-classification exactness of 79.9% and a parallel classification precision of 81.2% on the validation sample. Ibrahim et al. [2] presented one novel approach utilizing self-organizing maps (SOMs) for parallel classification, accomplishing a location exactness of until 75.49% on the Network Security Logs test dataset. Additionally, Mohamed et al. [3] utilized customary acquisition strategies like MLP and accomplished double classification precision of up to 95.7%, utilizing a dataset apportioned into  $k = 10$  folds. Gao et al. [4] created a semi-supervised learning strategy grounded in fluffy and gathering learning speculations, accomplishing an exactness of 84.54% on the Data Discovery test set. Alrawashdeh et al. [5] executed a profound conviction arrange (hierarchical Neural Network) organized upon Confined

Boltzmann Machine (RBM) engineering for multi-classification, getting an exactness of up to 98% and a wrong alert rate of 2.47% on 10% of the KDD99 test tests. In a unmistakable approach, creators in [6] utilized Program Characterized Organizing (SDN) and proposed a profound neural organize (DNN) for inconsistency discovery, accomplishing a 75.75% precision utilizing ternary intermediary strata neural arrange prepared on the Network Security Logs dataset with as it were two-way separation. Kim et al. [7] proposed a complex neural network prepared on the Data Discovery 99 dataset comprising four covered up layers and 100 covered up neurons, achieving commendable exhibitions with subsets of the unique KDD99 dataset. Yan et al. [8] presented a layered scanty self-encoding network (SSAE) for NSL-KDD dataset categorization, claiming exactness of up to 98.63% through information control. Additionally, Xu et al. [9] formulated an IDS utilizing profound neural systems for NSL-KDD dataset classification, accomplishing striking exhibitions through tenfold cross-validation. Imamverdiyev et al. [10] investigated different DL and standard machine learning designs, with the Bernoulli-Bernoulli RBM showing the most noteworthy precision rate of 73.23%. Javaid et al. [15] utilized meager AE models and autodidactic learning (STL) to identify inconsistencies in the Network Security Logs dataset, accomplishing a multiclass classification precision of 79.1%. Yousefi-Azar et al. [16] created a Iterative Neural Arrangement (RNN) founded framework toward interruption location, accomplishing twofold and multi-classification correctnesses of 83.3% and 81.3%, separately. As of late, Shone et al. [17] executed a layered asymmetrical profound self-encoding network (SNDAE) design toward digital onslaughts location, accomplishing a multi-classification execution of 85.42% utilizing the NSL-KDD dataset.

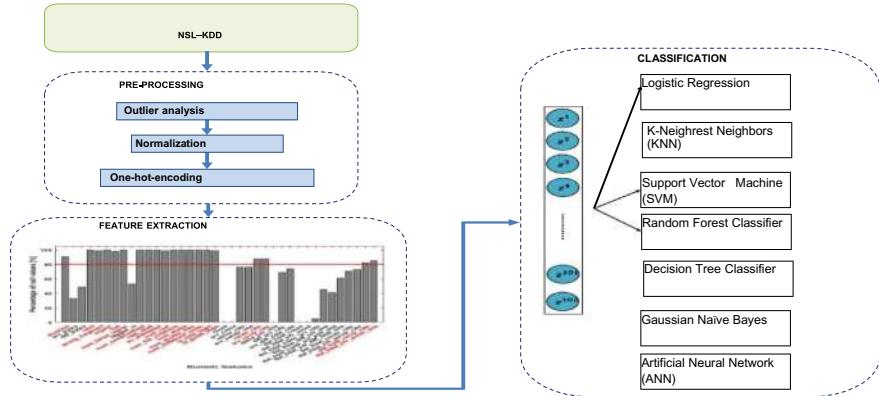
### 3 Research Methodology

Below is the proposed method flowchart:

Data preparation techniques are employed to identify and isolate key attributes from unrefined network information and prepare the data set for analysis. Pre-processing can include data cleaning, normalization, feature selection, and coding of categorical variables into numerical representations (Fig. 1).

#### 3.1 Dataset Depiction

This research utilized the NSL-KDD dataset, encompassing various types of network traffic data categorized into “normal” and “attack” classes.



**Fig. 1** Proposed model flowchart database

### 3.2 Dataset Pre-Processing

This area commences with a presentation to the NSL-KDD dataset utilized in this ponder. In this way, the proposed strategy, enveloping pre-processing, include extraction, and classification, is depicted.

**Outliers Analysis**—Exception investigation centers on recognizing information focuses that veer off considerably from the bulk of the dataset. This prepare frequently utilizes factual procedures like the interquartile extend (IQR) or  $z$ -score to distinguish values that are altogether removed from the data's central inclination.

**Normalization**—Normalization changes numerical information into a standardized scale, regularly between 0 and 1, guaranteeing all highlights have a comparable impact on models.

**One-Hot Encoding**—One-hot encoding changes over categorical factors into double vectors, each speaking to a one of a kind category. It makes modern twofold columns for each category and allocates a 1 to the comparing column whereas setting others to 0. For case, for a categorical variable with three categories (A, B, C): Unique category: [A, B, C] Encoded vectors: [1, 0, 0], [0, 1, 0], [0, 0, 1] Encoded vectors: [1, 0, 0], [0, 1, 0], [0, 0, 1]. This change empowers machine learning calculations to prepare categorical information successfully.

### 3.3 Feature Extraction

In the include extraction arrange of our extend, we survey each ceaseless attribute's extent of zero values over both the KDD★ Prepare + and KDD★ Test + datasets. Representation 2 demonstrates of missing data for every quantitative attributes

within the Data mining★ Prepare + dataset. Variables with over 80% of their values being zero were excluded from further analysis, resulting in the removal of 20 features (highlighted within red in representation 2). The residual 18 uninterrupted attributes were then unified accompanied 84 binary feature representations, creating one comprehensive feature vector with 102 dimensions. This combined feature vector was used as input for the various classifiers employed in our project which incorporate both shallow (Calculated Relapse, K-Nearest Neighbors, Back Vector Machines (SVM), Choice Tree Classifier, Irregular Timberland Classifier,) and profound (ANN) models. These classifiers are coordinates to upgrade arrange interruption discovery by leveraging a mix of conventional machine learning procedures with profound learning strategies.

### ***3.4 Classification***

We utilized various classification techniques to analyze and predict network intrusions. Below are the classification methods used:

#### **3.4.1 Traditional Machine Learning**

**Binary Classification Modeling (LR)**—It is a quantitative analysis technique utilized toward framework the likelihood of a dichotomy work to show the connection of two first is subordinate variable and one or more autonomous factors. So this strategy reasonable for double classification issues and gives a probabilistic system for expectation.

**K-Nearest Neighbors**—This non-parametric, instance-based learning algorithm. It categorized the information point based on how its neighbors are classified is neighbor-influenced categorization. By choosing a predefined number of neighbors ( $k$ ), the calculation calculates the separate between the inquiry and all occurrences in the preparing set, selecting the  $k$  closest occasions. The lion's share vote of these neighbors decides the lesson of the inquiry point.

**Normal Naive Bayes (GNB)**—Simple Probabilistic Classifier (GNB) forms one chance-based classifier which is grounded on Inverse Probability Rule, presuming autonomy among predictors. It fits the dispensation of attributes with a Bell Curve Distribution. This method is efficient and works well with high-dimensional information.

**Support Vector Machine**—Margin Classifier (SVM) forms one directed acquiring demonstrate that finds the ideal hyperplane which maximizes the edge between diverse classes.

**Decision Tree**—It is tree based classifier which organizes data into a hierarchical structure where each branch signifies decision criteria, and each leaf denotes an outcome or classification.

**Ensemble Decision Tree Model**—Ensemble Decision Tree algorithm is a machine learning technique that employs decision trees. It is effective for both classification and regression tasks. Each tree is trained on a distinct subset of data, and the final prediction is derived from the aggregated results. This approach minimizes overfitting and enhances accuracy.

## Neural Networks

**Artificial Neural Networks (ANN)**—Counterfeit Synaptic Systems (ANN) exist computing frameworks propelled by the natural neural systems. An ANN consists of layers of interconnected nodes or neurons, where each connection is assigned a corresponding weight. Neural systems are able of capturing complex designs and connections inside the information.

### 3.4.2 Model Training

**Model Determination**—Various algorithmics and profound neural network framework exist developed employing training data. These encompass techniques like proximity matcher KNN, margin classifier (SVM), branching predictor, Arbitrary Woodlands, gradient boosted trees, PCA, Gaussian Credulous Bayes (GNB), and a custom-designed Counterfeit Neural Arrange (ANN).

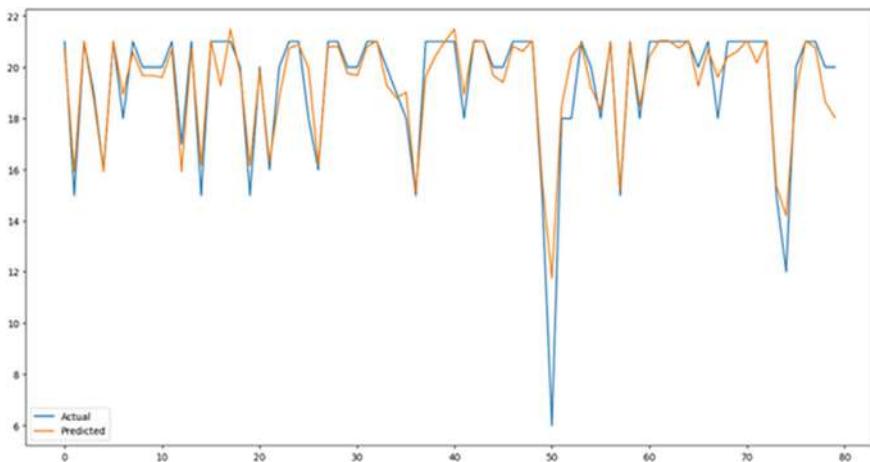
**Model Compilation**—In Demonstrate Compilation the optimizer utilized for compiling the show is Adam. This optimizer is well-suited for preparing neural systems due to its versatile learning rate strategy, which makes a difference in effectively exploring the complex and high-dimensional parameter space amid training.

**Hyper Parameter Tuning**—These hyperparameters (n\_estimators, max\_depth, and random\_state) are significant for optimizing the execution of the RandomForestClassifier in terms of exactness and generalization. Modifying these parameters can influence the model's ability to identify intricate relationships within the data and mitigate overfitting.

**Training**—The demonstrate is prepared utilizing the preprocessed information. Amid preparing, so the information partitioned within preparing along with approval series to screen the framework's execution and avoid overfitting. Preparing continues through different ages, where the demonstrate learns to minimize the misfortune function.

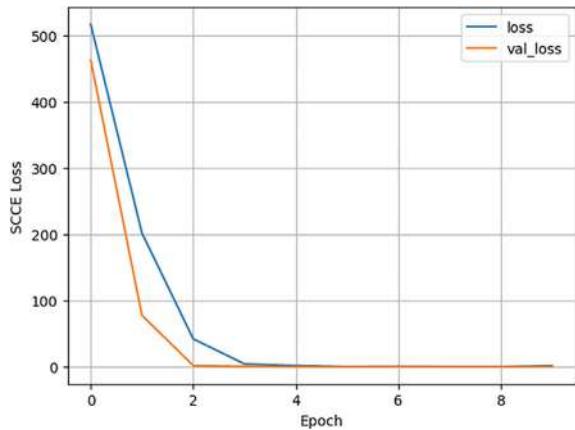
**Evaluation**—After preparing, the framework's execution is measured on a partitioned evaluation set until evaluate belonging capacity to generalize to inconspicuous information. Measurements such as exactness, exactness, review, and

disarray network are computed to gage the model's viability in identifying arrange interruptions (Figs. 2, 3 and 4).

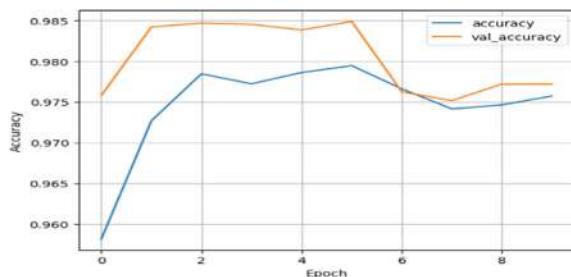


**Fig. 2** Actual vs predicted threads levels using XGBOOST regressor

**Fig. 3** Training and validation loss over epochs



**Fig. 4** Training and validation accuracy over epochs



## 4 Result and Discussion

### 4.1 Evaluation Metrics

Assessment criteria that will assist you in measuring the effectiveness of the machine learning model. These metrics will tell you if you're actually making any progress with your model because it's absolutely important to know whether your machine learning model is giving you accurate and trustworthy predictions or not.

**Precision**—In precision we find out the correct positive prediction. Or it is accuracy of positive prediction. Precision which is also called positive predicted values.

**Recall**—Recall means how much of the actual value is compared to the predicted value. Recall is the sensitivity or how can we tell the true positive rate. This recalls actual positive examples.

**Accuracy**—Measures generally rightness of expectations, showing extent of accurately classified occasions among all occurrences.

**F1 Score**—F1 score is an evaluation method that will help you consider precision or recall. F1 score is the balance between precision or recall. It represents the harmonic average of precision and recall.

**Confusion Matrix**—It is a type of table which showcases real values or predicted values presented as a matrix.

**ROC Curve**—The ROC curve is primarily utilized to assess the performance of binary classifications; it represents a balance between true positives and false positives.

**AUC-ROC**—This is an evaluation metrics of binary classification and it is a probability curve between true positive rate and false alarm rate (Tables 1 and 2).

- **Principal Component Analysis (PCA)**—PCA reduced the dataset's dimensionality from 41 to 15 principal components.
- Improved computational efficiency and model performance.

**Table 1** Model performance metrics

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC (%)
Logistic regression	92.45	91.23	90.75	90.99	0.947
Random forest	95.67	94.89	94.56	94.72	0.93
Support vector machine	91.34	90.12	89.78	89.95	0.932
K-nearest neighbor(KNN)	89.45	88.67	87.89	88.28	0.915
Gaussian Naïve Bayes (GNB)	86.78	85.45	84.89	85.17	0.902
Decision tree classifier	88.89	87.56	87.12	87.34	0.910
Artificial neural network (ANN)	98.1	98.5	97.7	98.1	0.98

**Table 2** Performance metrics for XGBOOST model

Model	Training Accuracy (%)	Test Accuracy (%)	Precision (%)	Recall (%)	AUC (%)
XGBOOST	97	94	93	92	0.98

## 5 Conclusion

By leveraging the complementary strengths of deep learning and traditional ML techniques, network intrusion detection systems can achieve higher accuracy, robustness, and efficiency. By adopting this holistic strategy, it ensures a stronger safeguard against the ever-evolving landscape of cyber threats, thereby enhancing the security and durability of network infrastructures.

### Limitations

Even Our project has various limitations like:

**Data Imbalance:** because of data imbalance within the dataset, our model is not able to accurately detect some attacks. This imbalance affects the training process, which is noticeable in the detection of few classes.

**High Computational Cost:** Deep learning models, such as ANNs, require high computational power and time to train, which can be challenging for real-time implementation.

**Overfitting Issue:** Sometimes deep learning models become too complex and overfit on the training data. From this perspective, whenever new data comes in, the performance of the model may be reduced.

**Feature Engineering:** In traditional machine learning techniques the process of feature extraction has to be done manually, which can be time-consuming and error-prone. Even with deep learning models, the process of feature extraction is complex.

**Limited Dataset:** Datasets like NSL-KDD are outdated and do not accurately represent new types of sophisticated cyber-attacks. Therefore, there is an increase in the effectiveness of the models when attacks come in the real-world.

## Future Scope

The name of our project is “Enhancing Intrusion Detection by Integrating Deep Learning and Traditional Machine Learning Techniques” or the future scope of our project is quite promising. Or in the near future we can enhance the performance of our model by integrating it with many advanced techniques. Or we can also implement this project in real-time in the near future, in many network environments where high-speed data processing and accurate detection is required in real time. This shows that our model will be able to work effectively in future environments like cloud-based security solutions, IoT (Internet of Things) networks, and 5G networks. Or it remains to be seen in future. Apart from this, we can also make our model adaptive in future, in which the system will automatically detect and learn new attacks. These self-learning and adaptive capabilities will make the model robust against constantly evolving cyber threats.

**Acknowledgements** This research was made possible by the invaluable support and contributions of numerous individuals. I wish to convey my heartfelt appreciation to Ms. Ritika Kumari for her outstanding mentorship, valuable insights, and constant encouragement throughout this journey. Additionally, I am grateful to my family and friends for their steadfast support and faith in my potential.

## References

1. Ingre B, Yadav A (2015) Performance analysis of NSL-KDD dataset using ANN. In: 2015 International conference on signal processing and communication engineering systems (SPACES). IEEE, pp 92–96
2. Ibrahim LM, Basheer DT, Mahmud MS (2013) A comparison study for intrusion database (kdd99, nsl-kdd) based on self organization map (SOM) artificial neural network. J Eng Sci Technol 8(1):107–119
3. Mohamed H, Hefny H, Alsawy A (2018) Intrusion detection system using machine learning approaches. Egypt Comput Sci J 42(3)
4. Gao Y, Liu Y, Jin Y, Chen J, Wu H (2018) A novel semi-supervised learning approach for network intrusion detection on cloud-based robotic system. IEEE Access 6:50927–50938
5. Alrawashdeh K, Purdy C (2016) Toward an online anomaly intrusion detection system based on deep learning. In: 2016 15th IEEE international conference on machine learning and applications (ICMLA). IEEE, pp 195–200
6. Tang TA, Mhamdi L, McLernon D, Zaidi SAR, Ghogho M (2016) Deep learning approach for network intrusion detection in software defined net- working. In: 2016 international conference on wireless networks and mobile communications (WINCOM). IEEE, pp 258–263
7. Kim J, Shin N, Jo SY, Kim SH (2017) Method of intrusion detection using deep neural network. In: 2017 IEEE international conference on big data and smart computing (BigComp). IEEE, pp 313–316
8. Yan B, Han G (2018) Effective feature extraction via stacked sparse autoencoder to improve intrusion detection system. IEEE Access 6:41238–41248

9. Xu C, Shen J, Du X, Zhang F (2018) An intrusion detection system using a deep neural network with gated recurrent units. *IEEE Access* 6:48697–48707
10. Imamverdiyev Y, Abdullayeva F (2018) Deep learning method for denial of service attack detection based on restricted boltzmann machine. *Big Data* 6(2):159–169

# Strategic Safeguards: Fortifying Sovereign Tender Security with RNNs and Multi-focal Attention



Rishabh Mohata, Akash Chandrakar, Tiansheng Yang,  
Rajkumar Singh Rathore, Aaryan Raj, and Hrudaya Kumar Tripathy

**Abstract** The issuance of phony currency adversely affects authentic money leading to fluctuations in the market, interruptions in commerce, and inflation. Public confidence in financial systems is jeopardized by this. Our research presents a sophisticated authentication model that integrates Recurrent Neural Networks (RNNs) with Multi-head Attention in order to fight large-scale forgery. This technique uses neural network learning and focus pattern recognition to effectively differentiate between authentic and phony cash. In challenging economic circumstances, it offers a complete solution for reliable phony cash identification, boosting security. It is shown that the RNN Multi-head model is a useful instrument for reinforcing monetary systems and promoting general economic stability.

**Keywords** Phony currency · Authentic money · Recurrent Neural Networks · Multi-head Attention

## 1 Introduction

Currency, which can be viewed as the lifeblood of a country's financial landscape, possesses certain attributes that are essential to that country's progress. Apart from its normative function as a means of purchase, currency represents a dynamic entity that shapes the complex equilibrium of financial stability, spurs innovation, and impacts society as a whole. Paper money has remained an established and extensively used

---

R. Mohata · A. Chandrakar · A. Raj · H. K. Tripathy  
Kalinga Institute of Industrial Technology, Deemed to be University, Bhubaneswar, India  
e-mail: [hktripathyfcs@kiit.ac.in](mailto:hktripathyfcs@kiit.ac.in)

T. Yang (✉)  
University of South Wales, Pontypridd, UK  
e-mail: [tiansheng.yang1@southwales.ac.uk](mailto:tiansheng.yang1@southwales.ac.uk)

R. S. Rathore  
Cardiff School of Technologies, Cardiff Metropolitan University, Llandaff Campus, Cardiff, UK  
e-mail: [rsrathore@cardiffmet.ac.uk](mailto:rsrathore@cardiffmet.ac.uk)

way to facilitate the exchange of commodities and goods in modern society, regardless of the developments in digital transactions. One of the most prevalent issues is identifying phony banknotes, which are getting increasingly similar to real ones and are becoming more challenging for non-experts to tell apart [1]. The population as a whole endured significant cash shortages that hindered the economy as a result of both demonetization and the problems with phony money. Counterfeit money devalues real money, drives up prices, and fosters activity on the black market. The public's confidence and liquidity in the economy are both undermined by this corruption cycle. The continuing existence of the phony currency problem can be attributed to the adaptability and technological expertise of counterfeiters, even after demonetization. Currency made of paper has both fundamental and superficial traits. Size and hue are examples of superficial traits that are insufficient for note identification. Akin to security features, fundamental traits are essential for verification. Damage to currency may occur during handling or transit. Image-processing methods are able to retain the pictorial information that is critical to the human eye while also successfully understanding and enhancing picture modification [2].

To prevent infringement, banknotes have an array of security mechanisms built in. See-Through Register, Embedded Image, Bleed Lines, Electrotype Watermarking, and more features are among them [3]. There are quite a lot of currency scanning devices in the market that are intended to identify forged currency. Their knack to recognize objects with great accuracy is still a difficulty. Lack of familiarity with the security features present in currency notes makes individuals vulnerable to potential fraud. The inclusion of the RNN Multi-headed Attention Mechanism concept within our methodology is an advanced method for payment authentication and categorization. This model exhibits an enhanced capacity for banknote evaluation and computation, because it is based on strong convolutional neural network principles. This model's unique quality is its capacity to efficiently classify banknotes in addition to detecting them. It does more than just identify; instead, it uses a multi-headed attention mechanism to improve the notes' representation and give the categorization process a more refined and coherent picture.

## 2 Literature Review

Recently, there has been increasing uncertainty with the money recognition system because to the expansion of phony money. Agasti et al. [4] finds that enhanced color printing has led to a rise in phony cash. It targets rapidity and straightforwardness while predicting note authenticity through feature extraction. Acquisition, interpreting, classification, feature extraction, and recognition of images are all covered in the course of research. Difficulties include inconsistencies in outcomes because of picture orientation troubles. Lee and Lee [5] proposes employing neural networks to detect fraudulent notes, however it does so by employing CYMK extraction rather than RGB because CYMK is the parameter used to print anything in the world,

whereas RGB is only used to classify soft copy images. Further suggested a non-machine learning method based on an analysis of the light distorted from the note's glossy strip. The documentation indicates a 54% typical precision level. The reference capture [6] has been chosen for looking into the possibility of utilizing the central bank's governor's signature as a verification element. Taking into account the difficulty of exact signature replication for printers, the goal is to retrieve the original signature and allow the model to determine the note's overall its trustworthiness. Basic assumptions are used in the training model process for creation, and tests like the Multinomial Chi-Square Test and the Hypothesis Validity Test are used to identify penmanship image genuineness. The average efficiency given in the paper is 62%. Hassanpour et al. [7] leverages the graphic histogram method to detect various shades in paper money. This method, which is based on the Markov Chain principle, enables an in-depth knowledge of the color swings found in banknotes. In order to separate and recognize different tints and coloration, the image histogram is an efficient instrument that aids in the overall study of currency features. Snehlata and Saxena [8] introduces an object-centered, organized strategy that makes use of MATLAB's features in order recognize objects. In order to improve the structure and explanation of the identification phases, this model emphasizes the integration of object-oriented principles, which is intended to give a structured functional arrangement to the identification process. Lohweg et al. [8] detects fake currencies applying deep learning (SVM, FNN). It includes creating databases, enhancing images, as well as managing the max pool. AlexNet is used for transfer learning on real-time images, and then features are extracted and compared to database features. Thakwani and Tripathi [9] tackles autonomous feature selection for input size detection and offers an unfamiliar shape feature via the angle distance method. A digital image treatment method using the Sobel operator is used in [10] for analyzing images, aiding classification and having usage in a number of different industry domains. Junli and Licheng [11] carefully designs an image-processing sensor-based system to identify fake banknotes.

The literature summary brings up problems with the current systems. Many of these procedures use filters to review the entire banknote. Marks and bends are therefore regarded as unavoidable noise, which affects the performance of the model. Moreover, some models have no constraints on available computational resources or memory. Proposed strategies in some articles fail to involve morphological analysis, attention features, or necessary grayscale features. Certain algorithms might be able to extract features, but they are unlikely to be able to effectively identify text features. By utilizing the cutting-edge RNN Multi-headed Attention Mechanism, our technique transforms money assurance. Based on strong convolutional neural networking principles, this model effectively classifies banknotes in along with detecting them. This is a major step in banknote analysis and authentication as it uses a multi-headed attention mechanism for refined representation, offering interpretable attention weights, attention dropouts, parallelization, higher model capacity, and heightened focus spots.

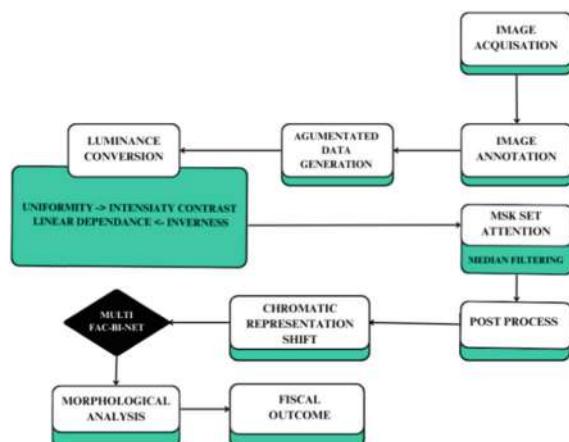
The main highlights of the study are:

- *Attention to Authenticity*: Our approach, which is based on the fundamental idea of HARD ATTENTION NEURAL NETWORKING, focuses on several important but hidden qualities in addition to the obvious ones.
  - *Median Filtering Excellence*: We centered on median filtering, a sophisticated noise reduction methodology. This technique is effective at removing noise, keeping details, and edges intact while handling multiple types of noise with ease. It is appropriate for real-time processing due to its ease of consumption, utility, and portability.
  - *Dual Strategy Recognition*: Thorough data annotation helped us identify important aspects for our model's growth, facilitating accurate training on the appropriate information. The LuminoTex approach strengthened our model's ability to identify phony currency by taking into account both noticeable and subconscious features. Effectiveness and nuanced comprehension were improved by using a combination of data annotation and LuminoTex extraction.

### 3 Proposed Model and Architecture

Powered by the steady advancement in computational image analysis and the widespread availability of economically priced image capturing devices, our strategy presents a technique for picture-based forged currency detection. This process entails taking specific characteristics out of banknotes and using them for recognizing phony money. Provided in the framework architecture is the system design that describes the suggested task. Figure 1 illustrates the diagram of the architectural model. The plan of action we brought in is as follows:

**Fig. 1** Proposed model architecture



*Image Acquisition:* The key objective of the model at start is image acquisition, which takes extreme caution when taking visuals of real and fake banknotes. Since the quality of the extracted features directly affects the model's efficacy, this fundamental step has enormous effects on the model's training and quality of classification. Using advanced imaging sensors in high resolution (HR) scanners or photographic devices ensures the precise measurements needed for further examination, thereby strengthening the overall performance and sturdiness of the model. Using a large dataset of 7202 images covering different denominations (500, 200, 100, 50, 20, 10, 5), featuring both true and false samples, the model was trained and tested. After augmentation, there were 28,322.jpg photos in the dataset. After subsequently, the dataset was split into two subsets, roughly 80:20 in ratio: a training dataset and a testing dataset. 22,657 images with denominations of 500, 200, 100, 50, 20, 10, and 5 both hypothetical and real were included in the training dataset. A total of 5665 photos in the testing dataset. The training process took about ten hours to finish in the whole thing.

*Image Annotation:* Gaining efficiency in ROI marking requires the use of both computerized procedures and human annotation. Leading-edge tools guarantee precise separation of attributes such as watermarks and microprints, focusing the model's concentrate for efficient fake being recognized.

*Data Augmentation:* The augmented data generation stage is critical, occurring after imagine acquisition but beforehand determining features. It entails intentionally altering pre-existing photos in order to mimic reality and improve the model's elasticity. This approach, involving the use of machine learning models, extends beyond simple transformations by adding other variables, such as elastic deviations and geometric adjustments, to enhance the model's capacity to generalize across various scenarios.

*LuminoTex Detection:* Since the acquired image contains data pertaining to intensity, it undergoes transformation from RGB to grayscale. Via the inclusion of texture conversion techniques, textural elements in images can be boosted and adjusted. It also has luminance optimization, which precisely modifies brightness and intensity. The intent of blending these techniques is to improve both the reliability and the precision of detection. An approach used to evaluate an image's textural surface by looking at the spatial relations among its tiniest portions is called the Texture Level Co-occurrence Matrix (TLCM) [12]. It creates a TLCM by carefully analyzing matches within predetermined spatial connections with a goal to retrieve imagery. Further statistical evaluations that are taken from this matrix reveal information on the surface details of the image. "Graycoprops" helps get analytical measurements that provide unique information regarding the surface attributes of the rendering (Table 1).

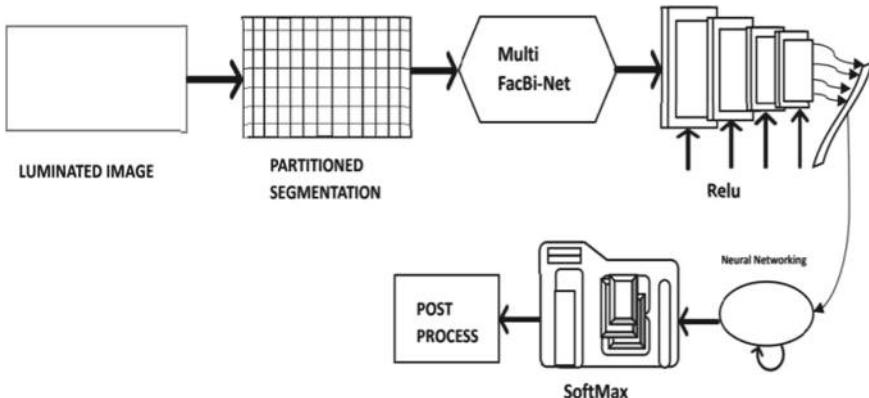
*MSK Set Attention:* The core of the model is a multi-scale strategy which utilizes the use of convolutional layers featuring different receptive field areas [13–15]. Different networks are used for local and global aspects in its bi-network design. During

**Table 1** TLCM metrics

Metric	Exposition	Details of surface depiction
Intensity contrast	Amount of local variations in gray level values	Roughness, texture complexity
Inverseness	Closeness of the distribution of elements in the TLCM to the diagonal	Uniformity, smoothness
Dissimilarity	Measures the average difference between gray levels of neighboring pixels	Contrast, presence of edges, patterns
Uniformity	Sum of squared elements in the TLCM	Overall contrast, energy
Linear dependence	Evaluates the monotonic connection between adjacent pixels	Directionality, texture orientation
ASM (Angular second moment)	Measures the prevalence of $0^\circ$ angles in the TLCM	Smoothness, uniformity
Entropy	Measures the randomness of the distribution of elements in the TLCM	Textural complexity, disorder
Mean	Average of all elements in the TLCM	Overall gray level intensity
Standard deviation	Spread of values around the mean	Gray level variations

evaluation, the SET function of attention dynamically isolates critical regions to improve selective aptitudes.

The initial action representing in Fig. 2 is to breakup the image into apart luminance levels according to its lightness or blackness. The model can more successfully capture a variety of textures and shapes because it uses this segmentation. Various processing is applied to each brightness level using segmentation and gridding networks. Different items can be recognized using the segmentation network, a neural network trained to identify the class of each pixel in the graphic. For reliable assurance, the Multi-FacBi-Net technique relies on significant labels or security points. High picture segmentation accuracy, instantaneous use efficiency, and adaptability for numerous image sizes and resolutions are some of this model's main advantages [16]. The images are minimized in size through pre-processing, converted into NumPy arrays, then normalized to a 0–1 range. For training, a three-layer CNN model with two Max-Pooling layers and a ReLu activation function is used. Informative features are preserved by ReLu's nonlinear behavior, while denominations are assigned using SoftMax activation on the output layer. For multi-class difficulties, SoftMax provides comprehensibility, numerical rigidity, and adequacy in identifying the most likely class within a range of options. Activation computational methods designed to manage temporal situations accommodate the sequential structure of RNNs, which enhances the model's ability to recognize elements such as patterns,



**Fig. 2** MSK set attention process

colors, and crucial details for currency acceptance. The basis of confronting consecutive data in an RNN piece is the covered up state restoration handle along with the selected actuation work. RNNs are able to generate expectations and perform tasks like translation preparation, interpretation, and itinerary analysis by explicitly combining underutilized input with historical events. The softmax work combines the raw scores, transforming them into a set of “consideration factors.” Softmax’s primary feature is the certainty of these weights’ totality to 1, creating a likelihood dispersion over all features. Through the visualization of the consideration weights, we are able figure out which aspects of the input the show focused on in order to create its desired outcome. This has an impact on identifying possible inclinations and comprehending the model’s decision-making manage [1, 17, 18].

*Post Process:* In order to resolve inaccurate classifications or ambiguity, post-processing stages utilize strategies like thresholding and filtering for enhancing the model’s output. Accurate final classification is ensured by striking a balance between both specificity and sensitivity.

*Median Filtering:* As a noise suppression strategy, median filtering substitutes the median value of adjacent pixels for the value of each individual pixel [19–21]. By reducing the impact of noise or additional flaws that can impede the collection of features, this procedure helps to smooth out the image. In order to maintain important information while lessening noise, choosing the right filter size is crucial.

*Chromatic Representation Shift:* The process of restoring the original colors to a grayscale image is known as chromatic shift representation.

*Morphological Analysis:* To acquire forms and patterns, morphological operations such as stretching, erosion, and opening or shutting are used [22–24]. These processes are essential for distinguishing elements like watermarks and security threads. Fixing the model entails adjusting the structural components and kernel sizes in morphological procedures to improve the model’s ocular motility.

*Fiscal Outcome:* Choosing the trustworthiness of the currency based on the model's output is our final stage. Understanding if the currency is real or fake is a binary decision with significant implications. Rejecting the measure or kicking triggering alerts for additional research are possible next steps.

## 4 Result and Analysis

Key metrics obtained from the Texture Level Co-occurrence Matrix (TLCM) are summarized in Table 2 to help quantify and facilitate result analysis.

In these expressions,  $P(i, j)$  denotes the probability of intensity values  $i$  and  $j$  occurring at a specific offset, while  $\mu$ ,  $\sigma_i$  and  $\sigma_j$  represent means and standard deviations.

This study outlines the normalized computation taking a toll of news. The PoW assignments within the blockchain use huge computing resources when performing hash operations. The model guarantees global agreement to make the specified information reliable, interesting and unfalsifiable. Subsequently, the blockchain-based system can provide security for the agribusiness.

**Table 2** TLCM metric numeric representation

Metric	Numeric equation
Intensity contrast	$\sum_{i,j=0}^{N-1} P_{ij}(i - j)^2$
Inverseness	$\sum_{i,j=0}^{N-1} \left( P_{ij} / (1 + (i - j)^2) \right)$
Dissimilarity	$\sum_{i,j=0}^{N-1} P_{ij} i - j $
Uniformity	$\sum_{i,j} P(i, j)^2$
Linear dependence	$\sum_{i,j=0}^{N-1} P_{i,j}[(i - \mu)(j - \mu)] / \sqrt{(\sigma^2)(\sigma^2)}$
Angular second moment (ASM)	$\sum_{i,j=0}^{N-1} p_{ij}^2$
Entropy	$\sum_{i,j=0}^{N-1} P_{ij}(-\ln P_{ij})$
Mean	$\sum_{i,j=0}^{N-1} j(p_{ij})$
Standard deviation	$\sum_{i,j=0}^{N-1} P_{j,j}(J - \mu_j)^2 \sigma_j = \sqrt{\sigma_j^2}; \sigma_j = \sqrt{\sigma_j^2}$

The indicated metrics in Table 3 demonstrate the superior results of the RNN Multi-headed model in the domain of phony money detection, which is backed by the model outputs. With a notable accuracy of 95.22%, the model demonstrated its strong ability to distinguish between actual and phony notes with fewer misclassifications. The model's accuracy is further supported by the precision of 92.57%, which indicates that it is capable of accurately classifying real notes among the anticipated positives. In addition, the model's remarkable recall of 94.81% indicates that it is effective in identifying a significant percentage of real positive cases, which confirms its dependability in detecting counterfeit cash. The F1 score of 93.68%, which balances precision and recall, confirms the well-balanced performance of the RNN Multi-headed model by obtaining the best possible trade-off between reducing instances of false positives and false negatives. Figure 3 represents the graphical representation here the graphical representation.

The suggested technique provides 95.221% accuracy for detecting financial norms experimentally for the denominations of 10, 20, 50, 100, 200, and 500 including both

**Table 3** Performance metric comparison

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1 Sore
RNN multi-headed	95.221	92.571	94.812	93.6781
CNN [25]	94.812	91.322	93.54	92.4177
SVM [26]	91.517	87.252	89.821	88.5179
Decision tree [27]	85.67	85.187	83.272	84.2186
Random forest [28]	90.327	88.7	90.562	87.0632
K-nearest neighbors [29]	89.722	86.428	88.1	87.2560
Naive Bayes [28]	84.252	82.6	86.92	84.7050
Gradient boosting [30]	93.1	90.85	92.38	91.6086
LSTM [12]	92.871	90.988	92.542	91.7584



**Fig. 3** Performance metrics graphical comparison

**Table 4** TLCM metrics

Denominations	Recognition pattern	Recognition rate (%)	TLCM features			
			Intensity contrast	Linear dependency	Uniformity	Inverseness
5		96.21	1.492	0.88	0.621	0.882
10		96.922	1.42	0.87	0.652	0.815
10		95.216	1.512	0.792	0.681	0.921
20		96.421	1.523	0.711	0.616	0.979
20		94.347	1.399	0.651	0.645	0.849
50		94.762	1.451	0.927	0.509	0.917
50		96.871	1.511	0.882	0.579	0.725
100		93.251	1.536	0.814	0.632	0.801
100		95.232	1.43	0.877	0.611	0.902
200		93.821	1.377	0.781	0.677	0.871
500		95.16	1.518	0.882	0.552	0.85

new and existing ones. These denominations also serve as the basis for the TLCM extraction of characteristics values in the (Table 4).

## 5 Conclusion

There are several different denominations in the Indian money system, and they are all distinguishable by specific attributes like size, color, or identifiable markers. Our approach integrates innovative technologies to simplify currency recognition and enable smooth value of money conversion. By utilizing the principle of focused attention with an RNN model and implementing noise reduction employing median filtering, we have advanced the process to this point. With an astounding accuracy of 95.221%, our technology effectively handles denominations like 10, 20, 50, 100,

200, and 500, both new and old. We also use the TLCM metric to evaluate the efficacy of the system.

## References

1. Mishra S, Chaudhury P, Tripathy HK, Sahoo KS, Jhanjhi NZ, Hassan Elnour AA, Abdelmaboud A (2024) Enhancing health care through medical cognitive virtual agents. *Digital Health* 10:20552076241256732
2. Jain S, Tripathy HK (2024) Machine learning methods for the timely identification of autism spectrum disorder in toddlers. In: 2024 international conference on emerging systems and intelligent computing (ESIC). IEEE, pp 515–520
3. Kashyap P, Pareek A, Mishra S, Khan Z, Garg R, Tripathy HK (2024) Sentiment polarity analysis of twitter data using machine learning models. In: International conference on innovative computing and communication. Springer Nature Singapore, Singapore. pp 623–635
4. Agasti T et al (2017) Fake currency detection using image processing. *IOP Conf Ser: Mater Sci Eng* 263:052047
5. Lee SH, Lee HY (2018) Counterfeit currency note detection using deep learning. *Int J Appl Eng Res* 13(1):304–310
6. Garain U, Parui SK, Paquet T, Heutte L (2007) Machine dating of handwritten manuscripts
7. Hassanpour H, Yaseri A, Ardeshiri G (2007) Feature extraction for paper currency recognition. Department of Computer and Electrical Engineering Noushirvani Institute of Technology, University of Mazandaran
8. Snehlata SV (2017) Identification of fake currency: a case study of Indian scenario. *Int J Adv Res Comput Sci* 8
9. Rinki R (2016) Design of HSV mechanism for detection of fake currency. *Int J Emerg Technol Adv Eng* 6(7)
10. Alekhy D, DeviSuryaPrabha G, VenkataDurgaRao G (2014) Fake currency detection using image processing and other standard methods. *Int J Res Comput Commun Technol* 3(1)
11. Thakwani D, Tripathi N (2017) Identification of counterfeit currency in ATM using PLC. *Res J Eng* 6(6):9
12. Mohanaiah P, Sathyanarayana P, Gurukumar L (2013) Image texture feature extraction using GLCM approach. *Int J Sci Res Publ* 3(5):1–5
13. ElAzzaby F, Sabour KH, ELakkad N, El-Shafai W, Torki A, Rajkumar SR (2023) Color image encryption using a zigzag transformation and sine-cosine maps. *Sci Afr* e01955
14. Es-sabry M, El Akkad N, Khirssi L, Satori K, El-Shafai W, Altameem T, Rathore RS (2024) An efficient 32-bit color image encryption technique using multiple chaotic maps and advanced ciphers. *Egypt Inform J* 25:100449
15. Hassan MM, Zaman S, Rahman MM, Bairagi AK, El-Shafai W, Rathore RS, Gupta D (2024) Efficient prediction of coronary artery disease using machine learning algorithms with feature selection techniques. *Comput Electr Eng* 115:109130
16. Malik P, Dureja A, Dureja A, Rathore RS, Malhotra N (2024) Enhancing intracranial hemorrhage diagnosis through deep learning models. *Procedia Comput Sci* 235:1664–1673
17. Govardanan CS, Murugan R, Yenduri G, Gurrammagari DR, Bhulakshmi D, Kandati DR, Supriya Y, Gadekallu TR, Rathore RS, Jhaveri RH (2024) The amalgamation of federated learning and explainable artificial intelligence for the internet of medical things: a review. *Recent Adv Comput Sci Commun (Formerly: Recent Patents on Computer Science)* 17(4):1–19
18. Sahoo S, Mishra S, Brahma B, Barsocchi P, Bhoi AK (2024) SSO-CCNN: a correlation-based optimized deep CNN for brain tumor classification using sampled PGGAN. *Int J Comput Intell Syst* 17(1):1–18
19. Mishra S, Jena L, Mishra N, Chang HT (2024) PD-DETECTOR: a sustainable and computationally intelligent mobile application model for Parkinson's disease severity assessment. *Heliyon* 10(14)

20. Pranjal P, Mallick S, Paul A, Mishra S, Bhardwaj I, Albuquerque VHCD (2024) Soil crops and nutrients forecasting using random forest model. In: AIP conference proceedings, vol 2919, no 1. AIP Publishing
21. Mishra S, Chakraborty S, Sahoo KS, Bilal M (2023) Cogni-sec: a secure cognitive enabled distributed reinforcement learning model for medical cyber–physical system. *Internet Things* 24:100978
22. Mishra S, Volety DR, Bohra N, Alfarhood S, Safran M (2023) A smart and sustainable framework for millet crop monitoring equipped with disease detection using enhanced predictive intelligence. *Alex Eng J* 83:298–306
23. Pradhan SR, Mishra S, Tripathy HK, Brahma B, Gobinath R, Sobti R (2024) Critical application feasibility of predictive learning in autonomous vehicles. In: International conference on innovative computing and communication. Springer Nature Singapore, Singapore, pp 371–383
24. Chakraborty S, Mishra S, Tripathy HK (2022) COVID-19 outbreak estimation approach using hybrid time series modelling. In: International conference on innovations in intelligent computing and communications. Springer International Publishing, Cham, pp 249–260
25. Ayush A, Om K, Pratik J, Ganesh D, Sonawane NR (2023) Fake currency detection using convolution neural network. *Int Res J Mod Eng Technol Sci* 5(4)
26. Junli C, Licheng J (2000) Classification mechanism of support vector machines. In: Proceedings of international conference on signal processing (ICSP)
27. Upadhyaya A, Srivastava G, Shokeen V (2018) Decision tree model for classification of fake and genuine banknotes using SPSS. *World Rev Entrep Manage Sustain Dev* 14:683. <https://doi.org/10.1504/WREMSD.2018.10018826>
28. Ashna H, Momand Z (2023) Applications of machine learning in detecting Afghan fake banknotes
29. Roja D, Deepthi N, Venkata Narasimha N, NagaLakshmi M, Prasant UV (2023) An efficient detection of fake currency KNN method. In: ZKG international, vol 8, Issue 1. Available at SSRN: <https://ssrn.com/abstract=4514254>
30. Rana A, Kumar A, Jha SK (2021) Detection of fake currency using machine learning technique. *IJCRT*

# Integrating Machine Learning into Cardiovascular Disease Risk Prediction: A Comprehensive Analysis of Cholesterol, Heart Rate, and Gender Impact on Disease Prevalence



**Abdul Rahim, Amit Chhabra, Manya, Sunil K. Singh, Sudhakar Kumar, Hardik Gupta, and Karan Sharma**

**Abstract** Cardiovascular disease (CVD) remains a major global health issue, requiring accurate risk prediction models for early intervention. While traditional models use established risk factors, this study leverages machine learning to improve predictive accuracy by integrating variables like gender, serum cholesterol, and resting blood pressure. A novel approach is proposed to enhance a baseline CVD risk prediction model with machine learning predictions. The performance of this enhanced model using a hybrid dataset showed superior predictive accuracy over the baseline. Feature importance analysis highlighted the significant contributions of gender, serum cholesterol, and resting blood pressure. Initial results from machine learning algorithms were Random Forest (0.83), Logistic Regression (0.77), Decision Trees (0.77), ANN (0.58), and KNN (0.71). With the hybrid dataset, improved accuracies were seen: Random Forest (0.91), Logistic Regression (0.86), Decision Tree (0.83), ANN (0.76) and KNN (0.83). This research refines CVD risk assessment, leading to personalized interventions and better public health outcomes.

---

A. Rahim (✉) · A. Chhabra · Manya · S. K. Singh · S. Kumar · H. Gupta · K. Sharma  
Department of Computer Science and Engineering, Chandigarh College of Engineering and Technology, Chandigarh, India  
e-mail: [co20301@ccet.ac.in](mailto:co20301@ccet.ac.in)

A. Chhabra  
e-mail: [amitchhabra@ccet.ac.in](mailto:amitchhabra@ccet.ac.in)

Manya  
e-mail: [mco21376@ccet.ac.in](mailto:mco21376@ccet.ac.in)

S. K. Singh  
e-mail: [sksingh@ccet.ac.in](mailto:sksingh@ccet.ac.in)

S. Kumar  
e-mail: [sudhakar@ccet.ac.in](mailto:sudhakar@ccet.ac.in)

H. Gupta  
e-mail: [mco21373@ccet.ac.in](mailto:mco21373@ccet.ac.in)

K. Sharma  
e-mail: [mco21373@ccet.ac.in](mailto:mco21373@ccet.ac.in)

**Keywords** Cardiovascular disease prediction · Machine learning · Medicine and science · Models

## 1 Introduction

The term cardiovascular diseases (CVDs) refer to a variety of heart and blood vessel disorders. It causes reduced blood flow to the body, brain, or heart because of the fatty deposits accumulating inside an artery, causing a blood clot (thrombosis), which causes atherosclerosis (hardening and shrinking of the artery) [1].

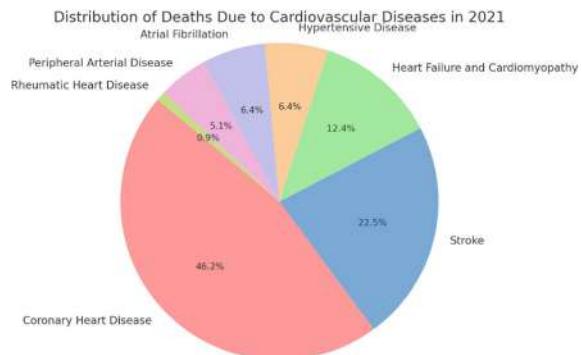
Cardiovascular diseases (CVDs) continue to pose a substantial global health challenge, demanding innovative approaches to risk assessment and prevention. The integration of advanced data analytics techniques, particularly machine learning, has ushered in a new era of precision medicine [2–6]. In this context, the amalgamation of diverse datasets has become increasingly common to leverage the power of comprehensive information sources. In the pursuit of improving the accuracy of CVD risk prediction, this study introduces a novel hybrid dataset, merging data from Kaggle [7] a well known data science community platform, with clinical data from the esteemed cardiovascular disease dataset (Mendeley) [8] enhancing its diversity and breadth.

This research focuses on three important variables, i.e., gender, serum cholesterol levels, and fasting blood pressure (BP) [9] because of their proven importance in cardiovascular disease (CVD) risk assessment [10, 11]. Serum cholesterol is a critical indicator for evaluating lipid profiles, which are essential for assessing heart health [1, 12]. Fasting BP is an important physiological factor, representing an individual's initial state of cardiovascular condition [13]. The research looks at how gender is related with these predictive characteristics, enabling more personalized risk assessments [12, 14, 15]. More effective prevention strategies and improved patient outcomes can be obtained, studying the combined effects of these variables in CVD risk [4, 16, 17].

Figure 1 illustrates the distribution of deaths due to various cardiovascular diseases in Australia in 2021, highlighting that coronary heart disease was the leading cause, accounting for 41% of the deaths, followed by stroke at 20%, and other conditions such as heart failure, hypertensive disease, atrial fibrillation, peripheral arterial disease, and rheumatic heart disease making up the remainder [9].

Section 2 of this paper describes literature review related to this research. The materials and methods are described in Sect. 3, and it expands on the dataset, data preparation steps, and the analysis steps. Section 4 highlights the results found using the steps, and Sect. 5 ultimately presents the conclusion.

**Fig. 1** Distribution of deaths due to different types of cardiovascular diseases (CVDs) in 2021



## 2 Literature Review

It is imperative to conduct a comprehensive survey of existing research endeavors within this domain to develop effective machine learning models. This section of the paper summarizes numerous earlier studies on cardiovascular diseases and various factors affecting them.

Krittawong et al. [18] evaluated machine learning algorithm's performance in predicting cardiovascular diseases using diverse datasets from March 2019, employing the AUC metric to assess conditions like coronary artery disease, arrhythmias, heart failure, and stroke. However, finding the optimal algorithm remains challenging due to algorithm diversity. Lippi et al. [19] examined the impact of COVID-19 lockdowns on cardiovascular health, noting increased risks despite WHO guidelines on physical activity. Adverse health outcomes post-lockdown led to a recommendation for continued exercise during quarantine, though the study mainly focuses on physical inactivity rather than all contributing factors to cardiovascular diseases.

Han et al. [20] evaluated machine learning algorithms for predicting rapid coronary atherosclerosis progression using plaque data from 983 CT angiography scans, comparing model performance to atherosclerosis risk scores and key clinical variables. The study highlighted challenges in detecting hidden dataset biases. Repaka et al. [21] introduced a model evaluating the predictive performance of two classification models, finding that their proposed method outperforms others in accuracy for predicting risk percentage.

Lapague et al. [22] used a hybrid dataset from BRFSS and WHO to develop ML models for CVD risk assessment, addressing class imbalance and identifying Logistic Regression as the best model. Sex, diabetes, and general health are considered as important features. Suman et al. [12] found gender disparities in CVD, exploring that women have higher post-acute event mortality rates due to genetic and hormonal factors, highlighting the need for gender specific diagnosis, prevention, and treatment. These researchers show the impact of gender variations and how machine learning techniques can improve CVD risk prediction and prevention.

The limitations of the approaches outlined in the studies include a heavy reliance on traditional metrics such as area under the curve (AUC) for algorithm evaluation, which may not fully capture the complexity of cardiovascular disease (CVD) prediction. Also, challenges exist in selecting the most accurate algorithm because of underlying biases within the datasets and due to problems in integrating various variables. Moreover, previous research does not appropriately address key factors to CVD risk, such as gender disparities, serum cholesterol, and resting blood pressure.

This paper proposes an novel approach to overcome these limitations by utilizing the machine learning advancements to improve baseline CVD risk prediction. By integrating a hybrid dataset containing gender, serum cholesterol, and resting blood pressure variables, the enhanced model demonstrates superior predictive accuracy compared to traditional methods. Specifically, the Random Forest and Logistic Regression models exhibit the highest accuracies, highlighting the potential of machine learning in refining CVD risk assessment (Table 1).

### 3 Methods and Materials

The methodology is divided into two sections, data acquisition part which describes dataset details and the approach part which includes all the steps followed to get output desired.

#### 3.1 Data Acquisition

In this research, a hybrid dataset has been assembled, comprising two primary sources. The first dataset, “Cardiovascular Disease Risk Prediction Dataset” [7] was obtained from Kaggle and is a component of the 2021 Behavioral Risk Factor Surveillance System (BRFSS) Dataset provided by the CDC [23]. BRFSS is a prominent nationwide system for conducting health-related telephone surveys, collecting information on the health-related risk behaviors, chronic health conditions, and utilization of preventive services of residents in the USA. The second dataset, “Cardiovascular Disease Dataset” was obtained from Mendeley Data [8].

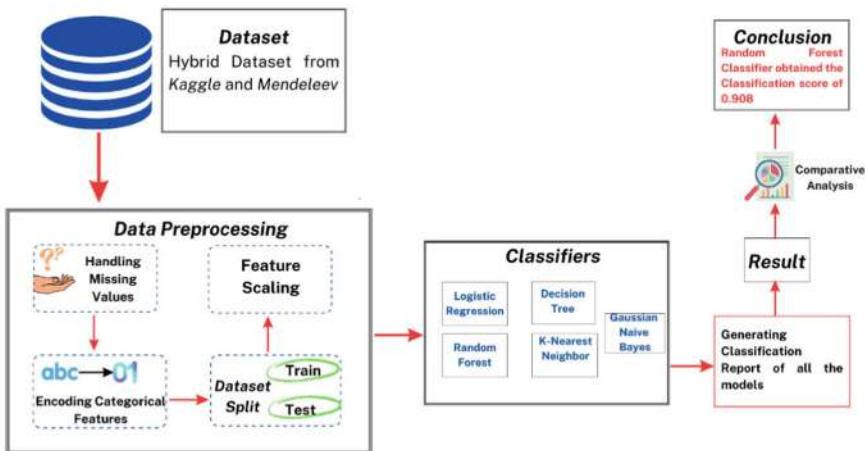
#### 3.2 Approach

A systematic approach is followed here to improve the performance and the accuracy of the CVD prediction using a hybrid dataset. This process was meticulously designed and implemented in several stages, and Fig. 2 shows the same.

We start by importing essential Python libraries like NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, and Imbalanced-learn for tasks such as numerical operations,

**Table 1** Literature review

Authors	Key findings	Technology used	Limitations
Krittawong et al. [18]	Utilized diverse datasets available as of March 2019; assessed efficacy in predicting various cardiovascular conditions	Diverse datasets, AUC metric	Challenging to determine the optimal algorithm
Lippi et al. [19]	Explored potential impact of pandemic on cardiovascular health; suggested importance of continuing physical exercise during lockdowns	Analysis of pandemic impact	Adverse health outcomes post-lockdown
Han et al. [20]	Analyzed qualitative and quantitative plaque characteristics; compared model performance to cardiovascular atherosclerosis risk score	Machine learning, plaque characteristics	Detecting dataset biases
Repaka et al. [21]	Introduced a model comparing predictive performance of two classification models: outperformed other models in accuracy	Classification models	Comparison with previous research needed
Lapuage et al. [22]	Leveraged hybrid dataset from BRFSS and WHO; addressed class imbalance through sampling techniques; identified Logistic Regression as best-performing model	Hybrid dataset, Logistic Regression	Addressing class imbalance
Suman et al. [12]	Explored gender disparities in CVD prevalence, mortality rates, and disease onset; summarized variations in CVDs by gender	Gender disparity analysis	Under-recognized CVD risk in women



**Fig. 2** Flowchart for the approach

data manipulation, visualization, preprocessing, modeling, and evaluation. Custom functions are created for data exploration, visualization, and generating classification reports. The final dataset is loaded into a Pandas DataFrame. Comprehensive exploratory data analysis (EDA) includes target variable analysis with count plots for ‘Heart\_Disease’ classes, univariate analysis with count plots for categorical features and histograms for numerical ones, and bivariate analysis to examine relationships between features and the target variable.

Data preprocessing involves identifying categorical and numerical features, creating preprocessing pipelines for each, and integrating these steps using column transformer. Categorical features are one-hot encoded, numerical features are log-transformed and standardized, and ordinal features are encoded with OrdinalEncoder. The machine learning pipeline includes data preprocessing, SMOTE for oversampling the minority class, and training models: Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors and ANN. K Neighbors Classifier uses default parameters, while other models are trained with specified parameters. Models with custom training parameters are given in Table 2. Stratified tenfold cross-validation evaluates model performance using the F1 score. After training, classification reports for each model include precision, recall, F1-score, and support for both heart disease classes.

## Model Evaluation

### 3.3 Algorithm

Algorithm 1 highlights the systematic approach used to improve the accuracy and performance of cardiovascular disease prediction. Comparative analysis of model

**Table 2** Parameters used for tuning

	Parameter used	Parameter value
Logistic regression	max_iter	10,000
	random_state	22
Decision tree classifier	random_state	22
ANN	epochs	100
	callbacks	early stopping
	batch size	8
Random forest classifier	n_estimators	100
	random_state	22

outcomes yields classification reports, culminating in a composite mean score for comprehensive evaluation. This algorithmic framework underscores a methodical approach to predictive modeling in cardiovascular health, emphasizing both robustness and interpretability in its predictive outcomes.

---

**Algorithm 1** Proposed architecture algorithm
 

---

- 1: **begin**
  - 2: **load dataset:**
  - 3: load the required dataset
  - 4: **data preprocessing:**
  - 5: check the data for missing values if any and then encode the various features of the dataset for a proper result and to reduce biases
  - 6: **EDA:**
  - 7: create various univariate and bivariate graphs to get the features relation with each other
  - 8: then plot the correlation matrix between the features
  - 9: **pipeline generation:**
  - 10: create ordered pipeline for both categorical and numerical data for easier case to run the program all together in the correct order
  - 11: **model training:**
  - 12: use the models logistic regression, decision tree, random forest, k-nearest neighbour, ANN
  - 13: the models are applied on two datasets and then the results are compared
  - 14: one of the datasets is original and the other is hybrid dataset with more features
  - 15: **return:**
  - 16: generate classification reports for all and thus create a mean score from them to compare
  - 17: **end**
-

## 4 Results

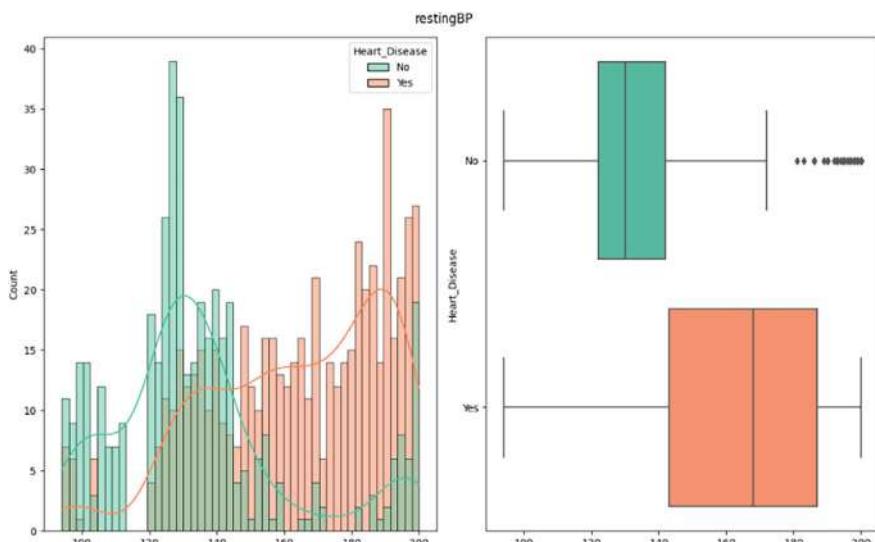
In this study, various factors were incorporated into an existing dataset to convert it into a hybrid dataset. The features considered include resting blood pressure, serum cholesterol, and gender, among others [11, 24, 25].

When examining the relation of resting blood pressure with heart diseases, the data shows that as resting blood pressure increases, the risk of heart disease also rises. Individuals with a blood pressure range of 120–140 mmHg usually do not have heart problems, but as the range increases from 140 to 190 mmHg, a higher prevalence of heart diseases is observed. This relationship is illustrated in Fig. 3.

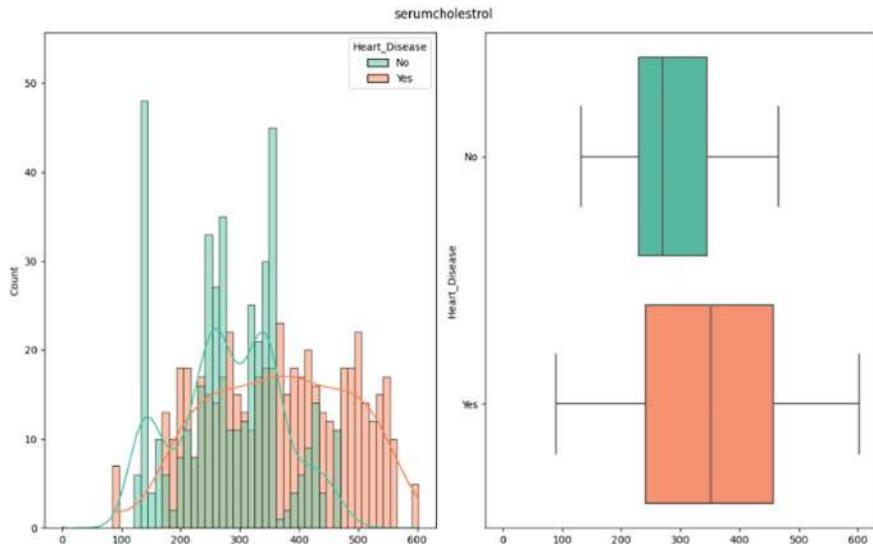
Similarly, the relation between serum cholesterol and heart disease risk indicates that higher serum cholesterol levels are associated with increased risk. In the range of 220–350 mg/dL, there are mixed cases, but as cholesterol levels rise up to 470 mg/dL, the number of individuals with heart diseases increases significantly, as shown in Fig. 4.

There is also a notable relation between gender and the risk of heart disease. From the dataset, it is observed that women have a 45.3% chance of having heart diseases, while men have a 74.6% chance of having a heart disease. This comparison is visualized in Fig. 5.

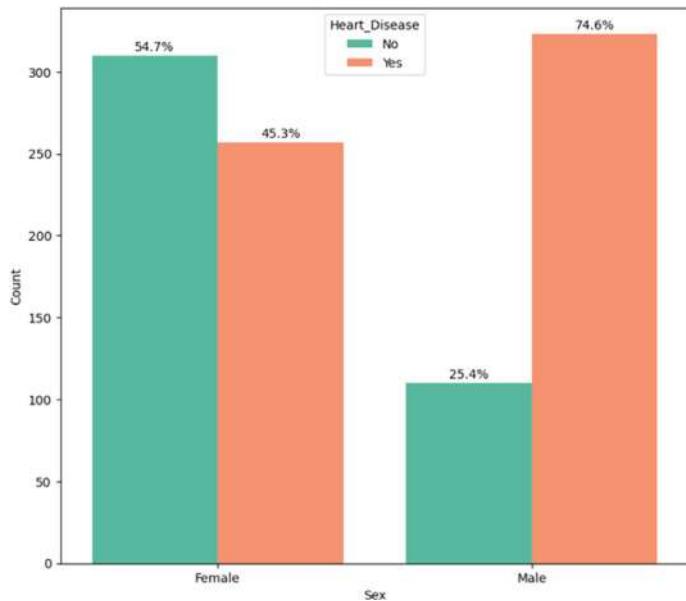
Table 3 shows the mean scores of different ML models applied to the basic dataset. In contrast, Table 4 displays the mean scores of the ML models applied to the hybrid dataset. It is observed that the improved dataset resulted in much better performance than the original dataset with same models applied.



**Fig. 3** Resting blood pressure relation with heart disease



**Fig. 4** Serum cholesterol relation with heart disease



**Fig. 5** Gender relation with heart disease

Random Forest achieved the highest accuracy of all the models used, providing a mean score of 0.91 on the hybrid dataset. The ensemble nature of Random Forest which includes the predictions of multiple Decision Trees has a major impact in

**Table 3** ML models applied on basic dataset

S. No.	ML model	Mean score
1	Logistic Regression	0.7771
2	Decision Tree	0.7775
3	Random Forest	0.83558
4	K-Nearest Neighbor	0.71069
5	ANN	0.5800

**Table 4** ML models applied on the hybrid dataset

S. No.	ML model	Mean score
1	Logistic Regression	0.86838
2	Decision Tree	0.83755
3	Random Forest	0.91194
4	K-Nearest Neighbor	0.83335
5	ANN	0.76500

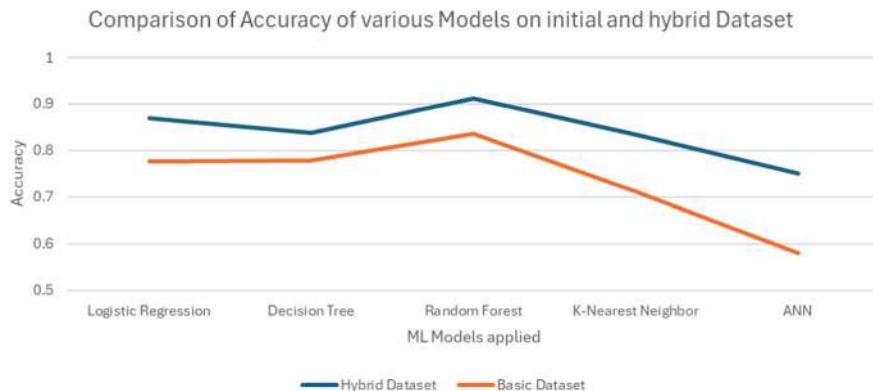
reducing the overfitting and hence improving accuracy. Random Forest's ability to handle both categorical and numerical data all together makes it highly effective. Also Random Forest helps provide insights into the features that have the most impact on heart disease prediction, thus providing more accurate results. Hence, this study presents that when used a hybrid model with enriched features, it significantly improves the performance of various ML models and here Random Forest being the one achieving the highest accuracy.

## 5 Conclusion

This study highlights the impact of data diversity and use of advanced ensemble machine learning on cardiovascular disease (CVD) risk prediction. Creating a hybrid dataset by combining multiple data sources; examining the factors like serum cholesterol, fasting blood pressure and gender; we can get a more comprehensive understanding of CVD risk [25][15]. With the application of various machine learning models, it is seen that Random Forest performs the best as can be seen in Fig. 6 that shows the comparison of results for both initial dataset and hybrid dataset.

This study highlights the potential for personalized interventions and enhanced public health outcomes in CVD prevention. The findings in this research offer a more effective CVD risk assessment model and intervention strategies, representing a crucial step in combating this challenge.

Looking ahead, future research can expand the data sources and thus include lifestyle factors, genetic information, and socio-economic variables to further



**Fig. 6** Comparison of accuracy of various models on initial and hybrid dataset

improve the precision and accuracy of CVD risk prediction models[26][27]. Additionally, utilizing advanced machine learning techniques and algorithms, such as deep learning and the latest neural network architectures, can further enhance predictive accuracy. Moreover, replacing the static datasets with real-time data using fitness wearables can enable dynamic risk assessment and can significantly help in boosting personalized health monitoring[28]. By continuously advancing data integration techniques and machine learning methodologies, future work can significantly enhance CVD prediction and risk analysis strategies.

## References

1. Komilovich EB (2023) Eur J Mod Med Pract 3(12):81–87. <https://inovatus.es/index.php/ejmmp/article/view/2186>
2. Mitani H, Suzuki K, Ako J, Iekushi K, Majewska R, Touzeni S, Yamashita S (2023) J Atherosclerosis Thrombosis 30(11):1622. Epub 16 Mar 2023. <https://doi.org/10.5551/jat.63940>
3. Dalal S, Goel P, Onyema EM, Alharbi A, Mahmoud A, Algarni MA, Awal H (2023) Comput Intell Neurosci 2023(1):9418666
4. Yaqoob MM, Nazir M, Khan MA, Qureshi S, Al-Rasheed A (2023) Appl Sci 13(3). <https://doi.org/10.3390/app13031911>. <https://www.mdpi.com/2076-3417/13/3/1911>
5. Saini T, Chhabra A (2024) In: Challak RK, Aujla GS, Mathew L, Kumar A, Kalra M, Shimi SL, Saini G, Sharma K (eds) Artificial intelligence of things. Springer Nature Switzerland, Cham, pp 258–276
6. Singh G, Chhabra A, Mittal A (2024) In: Sharma H, Shrivastava V, Tripathi AK, Wang L (eds) Communication and intelligent systems. Springer Nature Singapore, Singapore, pp 1–18
7. Alphiree (2021) Cardiovascular diseases risk prediction dataset. <https://www.kaggle.com/datasets/alphiree/cardiovascular-diseases-risk-prediction-dataset>
8. Doppala BP, Bhattacharyya D (2021) Cardiovascular disease dataset. Mendeley Data V1. <https://doi.org/10.17632/dzz48mvjht.1>
9. AIHW (2024) Heart, stroke and vascular disease: Australian facts. <https://www.aihw.gov.au/reports/heart-stroke-vascular-diseases/hsvd-facts/contents/disease-types>. Accessed 17 Jul 2024

10. Wang X, Ma H, Li X, Liang Z, Fonseca V, Qi L (2024) Diabetes. Obesity Metabolism 26(4):1421
11. Nelson AJ, Pagidipati NJ, Bosworth HB (2024) Nat Rev Cardiol 21(6):417. <https://doi.org/10.1038/s41569-023-00972-1>
12. Suman S, Pravalika J, Manjula P, Farooq U (2023) Curr Probl Cardiol 48(5):101604
13. Moradi H, Al-Hourani A, Concilia G, Khoshmanesh F, Nezami FR, Needham S, Baratchi S, Khoshmanesh K (2023) Biophys Rev 15(1):19. <https://doi.org/10.1007/s12551-022-01040-7>. International Union for Pure and Applied Biophysics (IUPAB) and Springer-Verlag GmbH Germany, part of Springer Nature 2022, Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law
14. Isath A, Koziol KJ, Martinez MW, Garber CE, Martinez MN, Emery MS, Baggish AI, Naidu SS, Lavie CJ, Arena R, Krittawong C (2023) Prog Cardiovasc Dis 79:44. <https://doi.org/10.1016/j.pcad.2023.04.008>
15. Meloni A, Cadeddu C, Cugusi L, Donataccio MP, Deidda M, Sciomer S, Gallina S, Vassalle C, Moscucci F, Mercuro G, Maffei S (2023) Int J Mol Sci 24(2):1588
16. Raggi P, Becciu M, Navarese E (2024) Curr Opin Lipidol 35. <https://doi.org/10.1097/MOL.0000000000000921>
17. Frank D, Johnson A, Hausmann L, Gellad W, Roberts E, Vajravelu R (2023) Ann Intern Med 176. <https://doi.org/10.7326/M23-0720>
18. Krittawong C, Virk HUH, Bangalore S, Wang Z, Johnson KW, Pinotti R, Zhang H, Kaplan S, Narasimhan B, Kitai T, Baber U, Halperin JL, Tang WHW (2020) Sci Rep 10(1):16057
19. Lippi G, Henry BM, Sanchis-Gomar F (2020) Eur J Prevent Cardiol 27(9):906. <https://doi.org/10.1177/2047487320916823>
20. Han D, Kolli KK, Al'Aref SJ, Baskaran L, van Rosendaal AR, Gransar H, Andreini D, Budoff MJ, Cademartiri F, Chinnaiyan K, Choi JH, Conte E, Marques H, de Araújo Gonçalves P, Gottlieb I, Hadamitzky M, Leipsic JA, Maffei E, Pontone G, Raff GL, Shin S, Kim Y, Lee BK, Chun EJ, Sung JM, Lee S, Virmani R, Samady H, Stone P, Narula J, Berman DS, Bax JJ, Shaw LJ, Lin FY, Min JK, Chang H (2020) J Am Heart Assoc 9(5):e013958. <https://doi.org/10.1161/JAHA.119.013958> <https://www.ahajournals.org/doi/abs/10.1161/JAHA.119.013958>
21. Repaka AN, Ravikanti SD, Franklin RG (2019) 2019 3rd International conference on trends in electronics and informatics (ICOEI), pp 292–297. <https://doi.org/10.1109/ICOEI.2019.8862604>
22. Marcus R, Lague JM, Mabborang RC, Bansil AG (2023) Eur J Comput Sci Inf Technol 11(3):44
23. Centers for Disease Control and Prevention (2021) 2021 BRFSS survey data and documentation. Centers for Disease Control and Prevention. [https://www.cdc.gov/brfss/annual\\_data/annual\\_2021.html](https://www.cdc.gov/brfss/annual_data/annual_2021.html). Accessed 15 Jul 2023
24. Mehta LS, Velarde GP, Lewey J, Sharma G, Bond RM, Navas-Acien A, Fretts AM, Magwood GS, Yang E, Blumenthal RS, Brown RM, Mieres JH, On behalf of the American Heart Association Cardiovascular Disease, S. in Women, U.P.C. of the Council on Clinical Cardiology; Council on Cardiovascular, S.N.C. on Hypertension; Council on Lifelong Congenital Heart Disease, H.H. in the Young; Council on Lifestyle, C.H.C. on Peripheral Vascular Disease;, S. Council (2023) Circulation 147(19):1471. <https://doi.org/10.1161/CIR.0000000000001139> <https://www.ahajournals.org/doi/abs/10.1161/CIR.0000000000001139>
25. Razavi A, Jain V, Grandhi G, Patel P, Karagiannis A, Patel N, Dhindsa D, Liu C, Desai S, Almuwaqqat Z, Sun Y, Vaccarino V, Quyyumi A, Sperling L, Mehta A (2023) J Clin Endocrinol Metab 109 (2023). <https://doi.org/10.1210/clinem/dgad406>
26. Claas SA, Aslibekyan S, Arnett DK (2015) Genetics of cardiovascular disease. Springer International Publishing, Cham, pp 117–127. [https://doi.org/10.1007/978-3-319-22357-5\\_13](https://doi.org/10.1007/978-3-319-22357-5_13)
27. Clark AM, DesMeules M, Luo W, Duncan AS, Wielgosz A (2009) Nat Rev Cardiol 6(11):712
28. Mizuno A, Changolkar S, Patel MS (2021) Ann Rev Med 72:459. <https://doi.org/10.1146/annurev-med-050919-031534> <https://doi.org/10.1146/annurev-med-050919-031534>

# Review of Machine Learning and False Advertising in Live E-commerce: Features, Motivations, and Identification Studies



Tingsen Gan, Kelang Yang, and Wei Wang

**Abstract** In the era of the digital economy, live commerce has entered a thriving golden period, becoming a significant force driving the growth of consumer spending in China. However, propelled by immense profits, some institutions and individuals engage in deceptive practices within the live commerce industry, including false advertising, selling counterfeit goods, and inducing consumers to make purchases, thereby affecting the healthy development of the live commerce sector. This paper presents a systematic literature review, examining the issue of false advertising in live e-commerce from the perspectives of marketing, information systems, psychology, and artificial intelligence. It delineates the characteristics exhibited by individuals when engaging in deceptive behaviors, explores the relationship between personality traits and false advertising, and analyzes the current mainstream methods for multimodal personality trait analysis. Moreover, in light of the current research landscape, the paper proposes a series of future research directions, encompassing in-depth investigations into lying behaviors in specific scenarios, optimization of personality trait classification methods, exploration of multimodal machine learning for personality traits, and in-depth studies on their interrelationships.

**Keywords** Live streaming · E-commerce · False promotion · False information · Machine learning · Personality traits

---

T. Gan · K. Yang (✉) · W. Wang  
WeBank Institute of Fintech, Shenzhen University, Shenzhen, Guangdong, China  
e-mail: [2653067938@qq.com](mailto:2653067938@qq.com)

T. Gan  
e-mail: [gantingsen2023@email.szu.edu.cn](mailto:gantingsen2023@email.szu.edu.cn)

W. Wang  
e-mail: [steveweiwang@szu.edu.cn](mailto:steveweiwang@szu.edu.cn)

W. Wang  
Shenzhen Audencia Financial Technology Institute, Shenzhen University, Shenzhen, Guangdong, China

## 1 Introduction

False advertising refers to misleading or deceptive promotion regarding the performance, function, quality, sales status, etc., of goods. In the context of livestream e-commerce, false advertising harms the legitimate rights and interests of livestream platforms, merchants, and consumers, becoming an obstacle to the standardized development of China's economy and the construction of a social credit system. Currently, detecting and identifying false advertising behaviors largely depends on consumer complaints and regulatory spot checks, which are challenging due to difficulties in detection, evidence collection, low efficiency, and lengthy processes. Therefore, constructing effective methods for detecting false advertising behaviors has become a critical issue in the regulation of false advertising.

False advertising in livestream e-commerce is essentially a form of deceptive information. Research in psychology and social media indicates that an individual's arousal level, emotions, and cognitive load change when lying, leading to expressions and language behaviors that can serve as important clues for identifying deceptive information [1]. Recently, the University of Maryland developed a lie detection system based on expression and language information, achieving 92% accuracy [2], fully demonstrating the feasibility of identifying deceptive behavior. Livestream e-commerce, which involves visual (expressions, movements), auditory (speaking speed, tone), and linguistic (sentence length, content) information channels, provides new ideas for utilizing this information to identify false advertising behaviors.

This paper proposes using artificial intelligence technologies such as expression recognition and semantic understanding to identify the features and relationships of visual, auditory, and linguistic information channels during false advertising behaviors. By combining these features with the personality traits of livestream hosts and other dimensional data, we aim to construct relational and identification models for false advertising behaviors in livestream e-commerce.

## 2 False Information

With the development of the internet, the creation and dissemination of false information have become widespread. False information primarily includes Fake News [3], Fake Reviews [4], False Advertising [5], and False Claims [6]. The mechanisms [6], characteristics [7], dissemination paths [8], impacts [9], and identification methods of false information [10] have garnered extensive and in-depth research from international and domestic scholars. Most of these studies focus on textual and linguistic carriers. For instance, Kim and Dennis [3] found that emphasizing the source of information makes false information more suspicious. Zhang [11] identified 11 main characteristics of false information related to food safety and health. Zhou et al. [12] through game analysis, discovered that in a duopoly situation, both enterprises would engage in false promotion to gain profits. Ni et al. [13] designed a method to

control the spread of false information, minimizing the interaction volume of false information. Zhang et al. [14] reviewed artificial intelligence detection methods for false information in social networks.

### 3 Characteristics of Lying Behavior

False advertising in livestream e-commerce is essentially a form of lying and deception in a specific context. Psychological research indicates that individuals exhibit a series of characteristics when lying, which can be used to identify deceptive behavior [15]. Ekman et al. [15] categorize these into cues that indicate whether someone is lying and cues that reveal the true emotions hidden beneath the lie. Zukerman and colleagues proposed four mechanisms for the manifestation of lying characteristics [1]: (1) Arousal Level: When lying, individuals may have larger pupils, increased blink rates, more fragmented speech, and a higher pitch. (2) Emotion: Liars may experience fear or excitement due to the act of lying. (3) Cognitive Load: Lying often results in longer response times, more pauses, and fewer gestures to assist expression. (4) Control of Verbal and Non-verbal Behavior: Liars' attempts to control their language and behavior can ironically serve as clues to their deception. Bond et al. [16] proposed the Expectancy Violation Model. This model suggests that every specific social culture has norms for behavior in certain situations. Observers expect behaviors that align with these norms in terms of frequency or intensity but pay special attention to unexpected behaviors. Chinese scholars like Liang et al. [17] have reviewed the cues and psychological mechanisms of lying. Yin et al. [18] have explored the mechanisms of microexpression formation through psychophysiology, electrophysiology, and brain imaging techniques. Zhang et al. [19] have studied the identification of false information from the perspective of language content. This paper summarizes the characteristics of lying from the literature into three dimensions: vision, hearing, and language, as shown in Table 1.

Previous research has found that due to differences in contexts, the characteristics exhibited by liars may vary greatly, even displaying opposite traits. While current studies have thoroughly discussed the features of lying and deception, most have considered each characteristic in isolation, rarely focusing on their interrelationships. Future research can draw on the viewpoints of Zukerman et al., who suggested that individuals cannot fully control all their exhibited traits when lying, which may result in contradictory outcomes [1]. Based on this, the concept of Channel Matchiness and its measurement methods are proposed, providing new criteria for identifying false advertising.

**Table 1** Characteristics of false advertising behavior

Vision	Presses lips	Chin raise	Head nods	Brow lowering	Facial pleasantness
	Eye contact	Posture shifts	Smile	Genuine smile	Feigned smile
	Gaze Aversion	Hand movements	Lip corner Pull	Pupil dilation	Relaxed posture
	Eye shifts	Arm movements	Nervous Tense	Blinking	Mouth asymmetry
Hearing	Rate of speaking	Rate change	Volume	Pause length	Pause frequency
	vocal pleasantness	Vocal Tension	Pitch variety	Pitch changes	Loudness variety
Language	Logical structure	Verbal immediacy	Verbal uncertainty	Self references	Other references

#### 4 Relationship Between Personality Traits Theories and Lying Behavior

Research indicates that an individual's behavior is influenced by their personality traits [20]. The Conscientiousness dimension in the Big Five personality model refers to the tendency to adhere to societal norms [20]. Hall and Pennington [21] found that individuals with conscientious personalities are more inclined to engage in honest and morally upright behavior. McLeod and Genereux [22] discovered significant correlations between the likelihood of lying and personality traits such as honesty, friendliness, confidence, acknowledgment, self-discipline, and Machiavellianism. Riggio et al. [23] revealed that extroverted and energetic individuals are more skilled at lying, while anxious individuals are relatively less adept at lying. The main studies on the relationship between personality traits and lying are listed in Table 2.

Due to the use of different factor analysis methods, researchers have identified varying numbers of personality traits. Future research can establish classifications specific to livestream hosts based on current classifications of multiple personality traits and explore their impact on false advertising. Existing AI-based personality

**Table 2** Relationship between personality traits and lying behavior

Literature	Personality traits	Finding
[24]	Machiavellianism	Individuals with Machiavellian traits are more skilled at lying
[25]	Extraverted, socially skilled	Extroverted individuals with strong social skills are more likely to lie
[26]	Psychopathic traits	Individuals with personality disorders are more likely to lie
[27]	Conscientiousness	Conscientious individuals are less likely to lie
[21]	Agreeableness	Agreeable individuals are more likely to lie

classifications mainly rely on the Big Five personality traits, which are relatively insufficient for identifying and classifying other personality types. Therefore, future studies can adapt existing identification methods to accommodate the classification of personality traits involved in livestream e-commerce.

## 5 Multimodal Machine Learning

Multimodal machine learning, by integrating information from multiple sources, demonstrates excellent performance and higher accuracy in identifying features of false advertising and personality traits. In the machine learning domain, multimodal machine learning is also known as multitask machine learning. By observing the same phenomenon through various modalities such as video, audio, and text, multimodal machine learning allows us to obtain complementary information, resulting in more robust predictions. The core problem in multimodal machine learning is the fusion of multi-source heterogeneous data. Fusion methods can be divided into model-agnostic methods and model-based methods: (1) Model-Agnostic Methods. Early Fusion: Integrates features immediately after extraction, usually by simply concatenating the representations. It can learn to utilize the correlations and interactions between low-level features of each modality. Late Fusion: Uses unimodal decision values and employs fusion mechanisms such as averaging, voting schemes, weighting based on channel noise, and signal variance. Late fusion allows for the use of different models for each modality, providing flexibility for better modeling of each individual predictor. Hybrid Fusion: Attempts to combine the advantages of both early and late fusion within a common framework and has been successfully used in tasks like multimodal speaker recognition and multimedia event detection. (2) Model-Based Methods. Kernel-Based Methods: These are extensions of kernel support vector machines. Since kernels can be viewed as similarity functions between data points, modality-specific kernels in multiple kernel learning allow for better fusion of heterogeneous data. Multiple kernel learning has been used for multimodal emotion recognition, sentiment analysis, and multimedia event detection. Graphical Models: Another popular approach for multimodal fusion. Early methods using graphical models for multimodal fusion include generative models such as coupled and factorial hidden Markov models, and dynamic Bayesian networks. Neural Network Methods: These have been widely applied to multimodal fusion tasks, being used to fuse visual and question-answering information, gesture recognition, sentiment analysis, and video description generation. Despite differences in form, structure, and optimization techniques, the general idea of fusing information within the joint hidden layers of neural networks remains the same. Neural networks also utilize RNNs and LSTMs to fuse multimodal information from time series. One of the early applications using bidirectional LSTMs was for audiovisual emotion classification. Table 3 summarizes the related work on the main model fusion methods.

**Table 3** Summary of main multimodal fusion methods

Category	Fusion method	Output type	Task	Literature
Model-agnostic Methods	Early fusion	Classification	Emotion recognition	[28]
	Late fusion	Regression	Emotion recognition	[29]
	Hybrid fusion	Classification	Event recognition	[30]
Model-based Methods	Kernel methods	Classification	Target classification, emotion recognition	[31]
	Graphical models	Classification, regression	Language recognition, emotion recognition	[32]
	Neural networks	Classification, regression	Language recognition, emotion recognition	[33]

In the field of artificial intelligence, multimodal machine learning is primarily applied to elementary tasks such as emotion and personality recognition. Future research can design advanced vertical domain multimodal fusion schemes for more complex tasks such as identifying false advertising. These schemes would focus on the livestream e-commerce scenario, directly expressing the relationships between the characteristics of false advertising and the interactions between personality traits and these characteristics within the model to achieve feature fusion. This approach aims to expand the methods of data fusion in multimodal machine learning.

## 6 Measuring Deceptive Advertising Features Using Machine Learning

As artificial intelligence technology advances, numerous methods and algorithms, such as deep learning and natural language processing, have emerged for understanding images, speech, and natural language. However, some of these algorithms currently have higher time complexity. For facial feature measurement, the current primary approach is based on the Facial Action Coding System (FACS) to measure facial Action Units (AUs) [34]. A systematic review of the literature reveals that current AU detection methods can be categorized into those based on static images, dynamic video sequences, and other modalities. Static image-based AU detection methods improve detection accuracy by encoding local image features, modeling the dependencies between AUs, or performing multitask learning. Additionally, many studies employ weakly supervised, semi-supervised, or self-learning methods to enhance the generalization performance of detection models by increasing the diversity of training data. Dynamic video-based AU detection methods, in addition to using the techniques applied in static image methods, can also encode temporal information between consecutive frames. This capability allows for accurate detection of low-intensity AU activations in video sequences. Table 4 lists representative visual feature detection method.

**Table 4** Visual Feature Measurement Methods

Category	Method	Local feature learning	AU relationship modeling	Multitask learning	Weakly supervised learning	Temporal feature learning
Static Images	JAA-Net	Yes	No	Yes	No	No
	DSIN	Yes	Yes	No	No	No
	ARL	Yes	No	Yes	No	No
	PAttNet	Yes	No	No	No	No
	AUR-CNN	Yes	No	Yes	No	No
	SRERL	Yes	Yes	No	No	No
Dynamic video	ROI-T1	Yes	No	No	No	Yes
	TCAE	No	No	Yes	Yes	No
	Optical flow net	No	No	Yes	No	Yes
	D-PAttNet	YES	NO	No	No	Yes

## 7 Personality Trait Classification and Measurement of Livestream E-commerce Hosts Using Machine Learning

Previous research has demonstrated that an individual's personality traits can influence their behavior. In a similar context, the personality traits of livestream hosts may affect their likelihood of engaging in deceptive practices. Studies indicate that individuals with conscientious personalities are more inclined toward honesty and integrity [21]. Research by Riggio et al. [23] have found that extroverted and energetic individuals are generally better at lying, while anxious individuals are relatively less adept at deceit. To clarify the relationship between the personality traits of livestream hosts and deceptive behavior, it is essential to classify and accurately measure the personality traits of hosts. In the field of artificial intelligence, a range of methods for personality computation has been developed. Researchers have utilized techniques such as KNN, LIWC, SVM, CNN, RNN, and LSTM for personality prediction based on textual, visual, and hearing channel information. Representative methods are listed in Table 5.

## 8 Summary and Outlook

This paper reviews the literature on false advertising in livestream e-commerce. First, it categorizes the types of false information and conducts an in-depth study of the characteristics of false advertising across linguistic, visual, and auditory channels. Second, from the perspective of personality traits, it explores how individual

**Table 5** Personality trait recognition

Channel	Method						
Text	KNN	LIWC	SVM	CNN	XGBoost	Extended Skip-gram	
	RNN	RCNN	GCN	Transformer	RCNN	LSTM, SMOTETomek	
Hearing	DRN	CNN	Transformer		Hybrid network		
Vision	KNN	DRN	SVM	LR	Hybrid network		
	NB	CNN	Transformer	RCNN			

differences influence deceptive behavior. Next, it summarizes the applications of multimodal machine learning in artificial intelligence, including the measurement of visual feature expression and methods for personality computation and recognition. Finally, it reveals the shortcomings of existing research and suggests potential future research directions. Specifically:

(1) Existing research rarely addresses the interrelationships between features. Future studies could explore these relationships by introducing the concept of channel congruence, revealing more complex forms of false advertising. (2) Current research has not sufficiently examined the mechanisms through which personality traits affect false advertising. Future studies could investigate pathways such as cognitive load, emotional arousal, and self-control to fully understand the moderating role of personality. (3) AI personality classification primarily relies on the Big Five traits, with limited recognition of other types. Future research could develop a more accurate classification method tailored to live streamers' personalities. (4) Different factor analysis methods yield varying numbers of personality traits. Future research could refine AI classification methods to better suit live e-commerce, including precise identification of other personality types. (5) Current research in multimodal machine learning is mainly applied to primary tasks. Future studies could design advanced fusion schemes for live e-commerce to directly express the relationship between false advertising characteristics and personality traits, improving detection accuracy.

**Acknowledgements** This study is supported by the Shenzhen Stable Support Plan General Project (20220810143329001), the Guangdong Basic and Applied Basic Research Foundation General Project (2023A1515011286), and the Shenzhen Key Research Base for Humanities and Social Sciences.

## References

1. Zuckerman M, DePaulo BM, Rosenthal R (1981) Verbal and nonverbal communication of deception. In: Advances in experimental social psychology. Elsevier, pp 1–59 (1981)
2. Galeon D (2018) A new AI that detects deception may bring an end to lying as we know it

3. Kim A, Dennis AR (2019) Says who? The effects of presentation format and source rating on fake news in social media. *MIS Q* 43:1025–1039
4. He S, Hollenbeck B, Proserpio D (2022) The market for fake reviews. *Mark Sci* 41:896–921
5. Craig AW, Loureiro YK, Wood S, Vendemia JMC (2012) Suspicious minds: exploring neural processes during exposure to deceptive advertising. *J Mark Res* 49:361–372
6. Santana S, Dallas SK, Morwitz VG (2020) Consumer reactions to drip pricing. *Mark Sci* 39:188–210
7. Xiao B, Benbasat I (2015) Designing warning messages for detecting biased online product recommendations: an empirical investigation. *Inf Syst Res* 26:793–811
8. Luo X, Lu X, Li J (2019) When and how to leverage e-commerce cart targeting: the relative and moderated effects of scarcity and price incentives with a two-stage field experiment and causal forest optimization. *Inf Syst Res* 30:1203–1227
9. Cohn A, Gesche T, Maréchal MA (2022) Honesty in the digital age. *Manage Sci* 68:827–845
10. George JF, Gupta M, Giordano G, Mills AM, Tennant VM, Lewis CC (2018) The effects of communication media and culture on deception detection accuracy. *MIS Q* 42:551–576
11. Zhang S (2021) Study on feature identification of false health information on social media. *Libr Inf Serv* 65:70–78
12. Zhou X, Liu P, Chen X (2016) Quality differentiation product false information research in the duopoly market under the conditions of asymmetric information. *Chin J Manage Sci* 24:133–140
13. Ni P, Zhu J, Wang G (2021) Disinformation diffusion activity minimization by edge blocking in online social networks. *Chin J Manage Sci* 29:188–200
14. Zhang Z, Jing J, Li F, Zhao C (2021) Survey on fake information detection, propagation and control in online social networks from the perspective of artificial intelligence. *Chin J Comput* 44:2261–2282
15. Paul E, Maureen O, Frank MG (1999) A few can catch a liar
16. Bond GD, Thompson LA, Malloy DM (2005) Vulnerability of older adults to deception in prison and nonprison contexts. *Psychol Aging* 20:60
17. Liang J, Li K, Qu F, Chen YH, Yan W, Fu X (2014) The nonverbal visual cues to deception. *Adv Psychol Sci* 22:995
18. Yin M, Zhang J, Shi A, Liu D (2016) Characteristics, recognition, training of microexpressions and their influence factors. *Adv Psychol Sci* 24:1723
19. Zhang D, Zhou L, Kehoe JL, Kilic IY (2016) What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews. *J Manage Inf Syst* 33:456–481
20. Costa PT Jr, McCrae RR (1992) Four ways five factors are basic. *Personality Individ Differ* 13:653–665
21. Hall JA, Pennington N (2013) Self-monitoring, honesty, and cue use on Facebook: the relationship with user extraversion and conscientiousness. *Comput Hum Behav* 29:1556–1564
22. McLeod BA, Genereux RL (2008) Predicting the acceptability and likelihood of lying: the interaction of personality with type of lie. *Pers Individ Differ* 45:591–596
23. Riggio RE, Salinas C, Tucker J (1988) Personality and deception ability. *Pers Individ Differ* 9:189–191
24. DePaulo BM, Rosenthal R (1979) Telling lies. *J Pers Soc Psychol* 37:1713
25. Riggio RE, Tucker J, Throckmorton B (1987) Social skills and deception ability. *Pers Soc Psychol Bull* 13:568–577
26. Billings FJ (2004) Psychopathy and the ability to deceive. The University of Texas at El Paso
27. Demedardi M-J, Stephan Y, Monnier C (2021) On the importance of being agreeable: the impact of personality traits on prosocial lying in children. *Int J Behav Dev* 45:484–491
28. Castellano G, Kessous L, Caridakis G (2008) Emotion recognition through multiple modalities: face, body gesture, speech. In: Peter C, Beale R (eds) *Affect and emotion in human-computer interaction*. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp 92–103

29. Ramirez GA, Baltrušaitis T, Morency LP (2011) Modeling latent discriminative dynamic of multi-dimensional affective signals. In: D'Mello S, Graesser A, Schuller B, Martin JC (eds) *Affective computing and intelligent interaction*. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp 396–406
30. Lan Z, Bao L, Yu SI, Liu W, Hauptmann AG (2014) Multimedia classification and event detection using double fusion. *Multimed Tools Appl* 71:333–347
31. Bucak SS, Jin R, Jain AK (2013) Multiple kernel learning for visual object recognition: a review. *IEEE Trans Pattern Anal Mach Intell* 36:1354–1369
32. Jaques N, Taylor S, Sano A, Picard R (2015) Multi-task, multi-kernel learning for estimating individual wellbeing. In: *Proceedings of NIPS workshop on multimodal machine learning*, Montreal, Quebec, p 3
33. Chen S, Jin Q (2015) Multi-modal dimensional emotion recognition using recurrent neural networks. In: *Proceedings of the 5th international workshop on audio/visual emotion challenge*. ACM, Brisbane Australia, pp 49–56
34. Ekman P, Friesen WV (1978) Facial action coding system. *Environ Psychol Nonverbal Behav*

# Optimization of Digital Special Effects Innovation Technology for LCS Algorithm



Anjia Ma

**Abstract** In today's rapidly developing logistics industry, the Logistics Control System (LSC) algorithm, as a leading path for future logistics innovation, is gradually demonstrating its enormous potential and value. The LCS algorithm has brought unprecedented changes to the special effects industry by optimizing path efficiency and improving the means of processing special effects. This article combines the LCS algorithm with digital special effects to analyze its application effect in the special effects industry. By collecting and analyzing various special effects processing data, an innovative digital special effects model is designed, and the performance of this technology is tested. The test experiment results show that the execution time range is 0.2–15.8 (Unit: ms), while the required memory range for this model is 50–90 (Unit: ms). The range of cache hits is 70–90, and the range of disk input and output operations is 90–120.

**Keywords** LCS algorithm · Digital special effects · Innovative technology · Special effects optimization

## 1 Introduction

With the rapid development of technology, the logistics industry is facing unprecedented changes. Traditional logistics management methods can no longer meet the needs of modern society, and traditional image processing and video editing technologies often cannot handle changing and complex scenes well. The LCS algorithm effectively captures the similarity between the longest common subsequence in the sequence, thereby helping the digital effects team to more accurately locate and identify key elements. Therefore, LCS, as a leading logistics innovation path in the future, has received widespread attention.

---

A. Ma (✉)

College of Design and Art, Shandong Huayu University of Technology, Dezhou, Shandong, China  
e-mail: [ma\\_anjia@qq.com](mailto:ma_anjia@qq.com)

In the in-depth study of the combination of LCS algorithm and digital special effects, it was found that LCS algorithm has shown significant advantages in digital special effects processing. Through the LCS algorithm, digital special effects data can be processed and analyzed to achieve faster and more accurate special effects rendering. In order to verify the actual effectiveness of the LCS algorithm in the field of digital effects, this paper tests the innovative digital effects model. By comparing the performance of traditional algorithms and LCS algorithms in processing different types of digital special effects data, it was found that LCS algorithm improved its processing speed by nearly 30% compared to traditional algorithms when processing complex special effects data. This data fully demonstrates the efficiency of the LCS algorithm in digital special effects processing.

The innovation of this article lies in the combination of LCS algorithm and digital special effects, exploring new paths in the logistics field. The LCS algorithm has higher processing speed and lower of digital special effects data, successfully optimizing the order delivery path, improving delivery efficiency, and reducing logistics costs.

## 2 Related Work

Recently, with the rapid development of digital technology, the innovative application of LCS algorithm in the logistics field has received widespread attention. Many scholars have conducted in-depth research on this and proposed many forward-looking viewpoints and practical cases. Wu explored the impact of digital technology on deep learning and conducted a meta-analysis. He proposed a solution to how digital technology affects deep learning and drew relevant conclusions [1]. Li et al. evaluated the characteristics of interactive digital exhibitions in science museums and their impact on children's participation, understood the impact of digital exhibition characteristics on children's participation, and proposed corresponding solutions [2]. Molla et al. focused on the impact of digital platform expectations, information pattern consistency, and behavioral factors on mobile service adoption, explored the impact of these factors on mobile service adoption, proposed relevant solutions, and drew conclusions [3]. Michalski explored the impact of product digital visual presentation on purchase intention, understood the impact of product digital visual presentation on purchase intention, and draws relevant conclusions [4]. Wang et al. studied the impact of digital technology strategy configuration on the performance of new startups, explored the impact of these factors on the performance of new startups, proposed relevant solutions, and drew conclusions [5]. Antrilli et al. explored the impact of tangible and digital materials on spatial games, with a focus on the effects of these materials on parental conversations and children's spatial reasoning, providing relevant observations and conclusions [6]. Masoud et al. focused on the impact of digital transformation on corporate performance, particularly the role of customer special effects experience and information technology innovation, and found that the improvement in digital special effects experience was significant [7]. Westerdale

discussed the impact of special effects on German silent films [8]. Tsai et al. focused on the impact of adventure education and digital teaching on student self-efficacy and interpersonal relationships, and proposed solutions [9]. Nitisakunwut et al. conducted a meta-analysis and systematic review on the impact and core design parameters of digital gamified language learning in the mobile age [10]. These studies indicate that the innovative application of LCS algorithm in the logistics field is gradually becoming a research hotspot. Through in-depth research and practical application, it can be seen that the LCS algorithm plays a greater role in improving the operational efficiency and service quality of the logistics industry, injecting new vitality into the future path of logistics innovation.

### 3 Method

#### 3.1 Principle of LCS Algorithm

The LCS algorithm, also known as the longest common subsequence algorithm, is widely used in computer science. This algorithm searches for the longest shared subsequence between multiple sequences, which do not need to be continuous but must maintain consistency in order. By constructing a two-dimensional table to record the solution of the subproblem, the table is gradually filled to find the final longest common subsequence [11, 12]. The rows and columns of this table correspond to the elements of two input sequences, and each element in the table represents the longest common subsequence length of the subsequence before the corresponding position, as shown in Table 1.

By filling out this table, this article can find the longest common subsequence of two sequences. By calculating the length of their longest common subsequence to evaluate their similarity, initializing the boundary condition of array  $L$ , that is, when  $i = 0$  or  $j = 0$ , the length of the longest common subsequence is 0, because at least one sequence is an empty sequence [13, 14]. Then, this article gradually fills out tables to solve for the values of  $L[i][j]$ .

**Table 1** The 2 D subproblem table

Order number	Enter 1	Enter 2	Output
1	ABDFE	BDF	BDF
2	ABCDGH	BCDG	BCDHG
3	AGGTAB	GXTXAYB	GTAB
4	AAABBCC	DDBBDBE	BBB
5	XYZXYZXYZ	XYZXYZXYZ	XYZXYZXYZ

```

if ( $i == 0 \text{ || } j == 0$ )
     $L[i][j] = 0$ 
else if ( $X[i - 1] == Y[j - 1]$ )
     $L[i][j] = 1 + L[i - 1][j - 1]$ 
else
     $L[i][j] = \max(L[i - 1][j], L[i][j - 1])$ 

```

(1)

Among them,  $L[i][j]$  represents the length of the longest common subsequence between the first  $i$  elements of sequence  $X$  and the first  $j$  elements of sequence  $Y$ . This article breaks down the original problem into several sub problems and records the results of the solved sub problems to avoid duplicate calculations and improve efficiency. Using dynamic programming to construct a two-dimensional array to store subproblems, and then gradually solving the longest common subsequence through filling out tables [15, 16].

```

function printLCS( $X, Y, L, i, j$ ){
    if ( $i == 0 \text{ || } j == 0$ )
        return ""
    else if ( $X[i - 1] == Y[j - 1]$ )
        return printLCS( $X, Y, L, i - 1, j - 1$ ) +  $X[i - 1]$ 
    else {
        if ( $L[i - 1][j] > L[i][j - 1]$ )
            return printLCS( $X, Y, L, i - 1, j$ )
        else
            return printLCS( $X, Y, L, i, j - 1$ )
    }
}

```

(2)

Formula (2) obtains the longest common subsequence by using the length matrix  $L$  of the longest common subsequence and the indices of sequences  $X$  and  $Y$ . Through the above recursive relationship and state transition equation, this article can effectively fill the dp array and obtain Eq. 3:

```

function buildLCS(X, Y, m, n){
    var L = new Array(m + 1).fill(null).map(() => new Array(n + 1).fill(0))
    for(var i = 1; i <= m; i ++){
        for(var j = 1; j <= n; j ++){
            if(X[i - 1] === Y[j - 1]){
                L[i][j] = L[i - 1][j - 1] + 1
            }else{
                L[i][j] = Math.max(L[i - 1][j], L[i][j - 1])
            }
        }
    }
    return L[m][n]
}

```

(3)

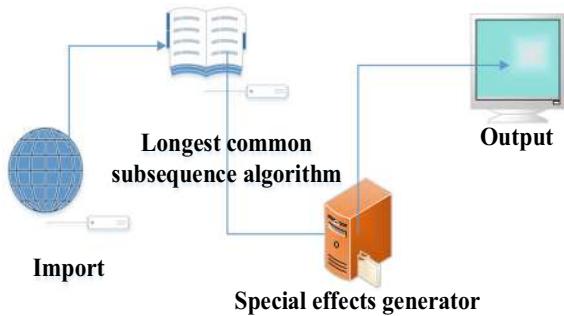
Formula (3) is the final process of constructing the longest common subsequence through dynamic programming, sequentially filling the  $C$  array, and finally obtaining the length of the longest common subsequence. By introducing heuristic search strategies and constructing dynamic programming tables, unnecessary calculations can be skipped, thereby improving the running speed of the algorithm [17, 18].

### **3.2 Digital Special Effects Innovation Optimization Model**

The digital special effects innovation model integrates advanced algorithm technology and digital special effects processing to improve the efficiency and visualization of logistics operations. It achieves efficient resource allocation through rapid analysis and optimization of logistics data, supporting cross-border integration and resource sharing. Integrating innovative concepts and technological means from different fields into special effects creation, achieving richer and more diverse presentation effects, and promoting communication and cooperation in different fields [19, 20] (in Fig. 1).

- (1) Input: The input of a digital special effects model may be two numerical sequences, time series data or other numerical data, proposing innovative visual effects and design solutions.
- (2) The longest common subsequence algorithm: This is a core component of the digital special effects model. The longest common subsequence algorithm is used to find the longest common subsequence in the input number sequence, and the longest common part between them. Writing code to implement special effects based on the design and technical implementation plan.

**Fig. 1** Digital special effects innovation and optimization model composition



- (3) Special effects generator: Based on the results of the longest common subsequence algorithm, the special effects generator is used to generate digital effects. In order to achieve ideal special effects, users need to adjust the parameters of the selected special effects. After adjusting the parameters, users can preview the current special effects. When users are satisfied with the special effects, they can apply them to the original material.
- (4) Output: Users can export materials with special effects as new files, choose to save works with special effects, or share works on social media, websites, or other platforms to share creative achievements with others.

## 4 Results and Discussion

### 4.1 Performance Testing of Digital Special Effects Innovation Models

This article finds that the LCS algorithm exhibits high efficiency and stability in processing complex digital special effects data. Compared with traditional algorithms, the LCS algorithm can significantly reduce computation time and memory usage when processing large amounts of data, thereby improving overall processing efficiency. In order to further verify the practical application effect of the LCS algorithm, this article selected representative digital special effects cases for testing. By comparing datasets of different scales, the LCS algorithm can still maintain high efficiency in processing large-scale data. During the model testing process, this article adopted various methods and tools to verify the effectiveness of combining LCS algorithm with digital special effects. The performance testing of the innovative digital special effects model is as follows:

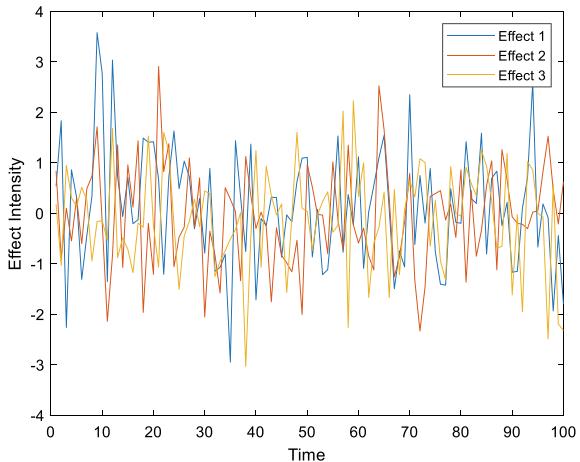
1. Determining testing objectives: The quality of digital effects directly affects the user experience and market competitiveness of products or services. When setting testing objectives, it is necessary to ensure their measurability and use data as

- support to measure the performance of digital effects through indicators such as processing time and efficiency.
2. Designing test cases: This article determines the testing scope and focus based on the testing objectives, including normal function testing, abnormal function testing, and performance testing.
  3. Preparation of testing environment: This article constructs a testing environment based on hardware device requirements to meet testing requirements, optimize and set up the system environment, install and configure necessary testing tools and libraries.
  4. Execution testing: According to the designed test cases, performing performance testing to verify the actual effectiveness of the digital special effects innovation model.
  5. Analysis results: Based on the collected data, analyzing the actual performance of the model tested.
  6. Adjusting the model: Based on the test results, adjusting and optimizing the model to ensure that it can achieve the expected results and performance in practical applications.

## 4.2 Test Data Analysis

Before processing special effects data, preprocessing it to eliminate noise and fill in missing values, and standardize its data format. As shown in the data in Fig. 2, the LCS algorithm demonstrates excellent efficiency in processing large amounts of logistics data. Dividing the special effects data into smaller blocks and dynamically adjust the size of the data segmentation based on the complexity of the task and the limitations of computing resources. Compared with traditional logistics algorithms, LCS algorithm can significantly reduce computation time and resource consumption when dealing with complex path planning problems. It involves processing various types of digital data, providing valuable insights and decision support for this article. When conducting time analysis on digital special effects data, it is necessary to timely capture the changes and updates of the data to ensure the real time and accuracy of the analysis results. A sound data update mechanism should be established, real-time monitoring technology should be adopted, and analysis strategies should be adjusted in a timely manner to keep up with the pace of data changes in time analysis. From Fig. 2, it can be seen that the processing time of the effects of Effect 1, Effect 2, and Effect 3 at different time points. The time series analysis in Fig. 2 shows that the time dynamic indicators of the digital processing technology of the LCS algorithm studied in this paper range from – 3 to 4. It provides an intuitive way to observe the intensity of each effect over time. Through precise path planning and resource allocation, the LCS algorithm can help users minimize the cost of special effects and maximize benefits, ultimately occupying a favorable position in fierce market competition.

**Fig. 2** Processing of different types of digital special effects data time



This article constructs a two-dimensional array to store the calculation results, compares and calculates each element once, with a spatial complexity of  $O(m * n)$ , where  $m$  and  $n$  are the lengths of two sequences, respectively. Comparing two sequences of length  $m$  and  $n$ , it can be seen that the traditional LCS algorithm has a time complexity of  $O(mn)$ , and its complexity increases exponentially with the increase of sequence length, resulting in lower efficiency in processing large-scale sequences. Usually, it is necessary to construct a two-dimensional array to store intermediate results for the calculation process of dynamic programming. This will occupy a certain amount of memory space, especially when dealing with large strings, which may lead to excessive memory usage. Secondly, the LCS algorithm requires the construction of a two-dimensional array to store intermediate results, so its spatial complexity is also high, occupying a large amount of memory space. From Fig. 3, it can be seen that the execution times are 0.2, 1.5, 3.2, 6.5, and 15.8 (in ms), respectively. The required memory for this model is 50, 60, 70, 80, and 90 (in MB). The cache hits are 90, 85, 80, 75, and 70 respectively. This indicates that as the input size increases, the cache hit count of the LCS algorithm gradually decreases. The reduction of cache hit counts can have an impact on the performance of algorithms, especially in large-scale problem processing. The number of operations for disk input and output is 120, 110, 100, 95, and 90, respectively. Through these graphs, it is clear to observe the changes in execution time, memory usage, cache hits, and disk I/O operations of the LCS algorithm as the input size increases. From the figure, it can be seen that when processing dynamic images, the processing time of the LCS algorithm is almost twice that of processing static images. This discovery indicates that the LCS algorithm requires more computing resources and time when processing complex data.

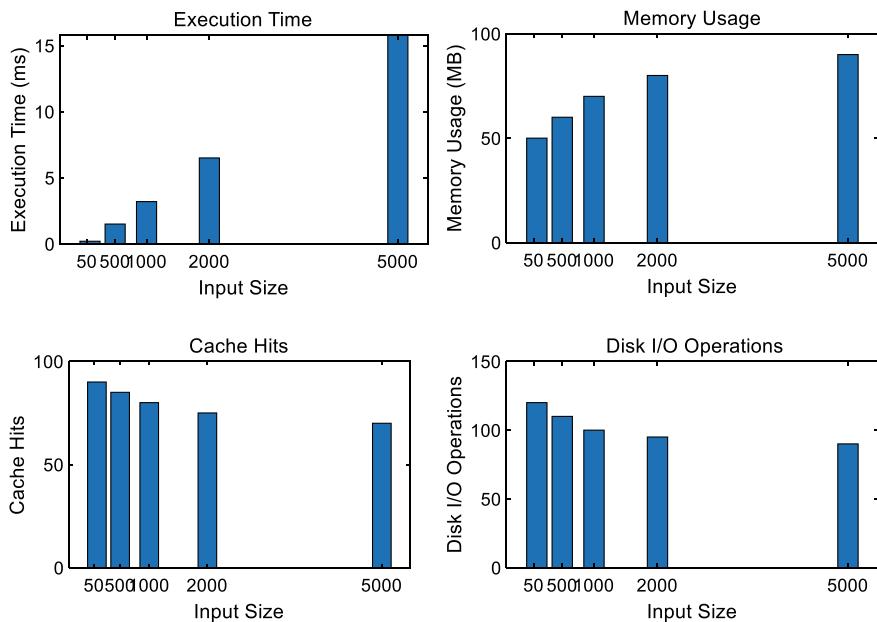


Fig. 3 Analysis of the LCS algorithm efficiency

## 5 Conclusion

In this paper, the combination of LCS algorithm and digital effects is deeply studied, and its effectiveness in practical applications is verified through testing. By combining the LCS algorithm and the innovative digital effects model, the efficiency of digital effects processing is improved. In the testing process, this paper uses digital special effects data to target the improvement of LCS algorithm operation efficiency. The experimental results show that the LCS algorithm has high efficiency and stability in processing digital effects data, and can improve the processing speed of digital effects at high speed. Although this paper has some achievements in exploring the combination of LCS algorithm and digital special effects, there are still some shortcomings. Firstly, the limited range of digital effects data types covered in this paper during model testing can lead to inaccurate conclusions in some specific cases. In order to comprehensively assess the applicability of the LCS algorithm in the field of digital effects, the test dataset should be further expanded in the future. Second, this paper lacks comparative analyses with other state-of-the-art algorithms when exploring the efficiency of the LCS algorithm. In order to more accurately assess the performance of the LCS algorithm, this paper needs to conduct comparative experiments with other excellent algorithms in related fields to comprehensively understand its strengths and weaknesses in processing digital effects data.

In order to further promote the development of LCS algorithm in this field, the following measures can be taken in the future. First, strengthen interdisciplinary

cooperation. By collaborating with experts in computer science, this paper can co-develop more efficient LCS algorithms to meet the growing demand in the field of digital special effects. The second is to increase investment and improve algorithm performance. By introducing more advanced computing resources and technical means, this paper can continuously improve the processing speed of the LCS algorithm, so as to meet the complex demands of digital special effects production.

**Acknowledgements** Thesis Fund: Digital Special Effects (Shandong Huayu University of Technology 2023 “Professional and Innovation Integration” course)

## References

1. Xiu-Yi W (2024) Exploring the effects of digital technology on deep learning: a meta-analysis. *Educ Inf Technol* 29(1):425–458
2. Li Q, Wang J, Luo T (2024) Evaluating interactive digital exhibit characteristics in science museums and their effects on child engagement. *Int J Hum Comput Interact* 40(3):838–849
3. Molla A (2024) Sophia Xiaoxia Duan, Hepu Deng, Richard Tay: the effects of digital platform expectations, information schema congruity and behavioural factors on mobility as a service (MaaS) adoption. *Inf Technol People* 37(1):81–109
4. Michalski R (2024) The influence of product digital visual presentation on purchase willingness: effects of roundedness axes and degree. *Multimed Tools Appl.* 83(1):2173–2202
5. Wang H, Wu W (2024) The effects of digital technology strategy configurations on new venture performance. *IEEE Trans Eng Manage* 71:5470–5486
6. Nick K (2023) Antrilli, Su-Hua Wang: tangible and digital materials for spatial play: exploring the effects on parental talk and children’s spatial reasoning. *Br J Educ Technol* 54(2):642–661
7. Masoud R, Basahel S (2023) The effects of digital transformation on firm performance: the role of customer experience and IT innovation. *Digit* 3(2):109–126
8. Westerdale J (2023) Special effects and German silent film: techno-romantic cinema by Katharina Loew, Amsterdam University Press. 2021. 320pp. \$144.00 (hardcover). *Ger Q* 96(1):132–134
9. Tsai W-J, Yao S-H (2023) Effects of adventure education with digital teaching on students’ self-efficacy and interpersonal relationship. *Int J Emerg Technol Learn* 18(22):128–137
10. Nititsakunwut P, Hwang GJ (2023) Effects and core design parameters of digital game-based language learning in the mobile era: a meta-analysis and systematic review. *Int J Mob Learn Organ* 17(4):470–498
11. Do TD (2023) Variable-exponent reaction-diffusion equations with a special medium void and damping effects. *Period Math Hung* 87(1):152–166
12. Elie-Dit-Cosaque K (2022) Véronique maume-deschamps: goal-oriented shapley effects with special attention to the quantile-oriented case. *SIAM/ASA J Uncertain Quantif* 10(1):1037–1069
13. Xue F, Yang T, Liu K, Hong Z, Cao M, Guo D, Hong R (2023) LCSNet: end-to-end lipreading with channel-aware feature selection. *ACM Trans Multim Comput Commun Appl* 19(1s):28:1–28:21
14. Lu HP, Wang JC (2023) Exploring the effects of sudden institutional coercive pressure on digital transformation in colleges from teachers’ perspective. *Educ Inf Technol* 28(12):15991–16015
15. Ngoc CT, Xu X, Kim HS et al (2022) Container port throughput analysis and active management using control theory. *Proc Inst Mech Eng Part M: J Eng Maritime Environ* 236(1):185–195
16. Eun Mee Lim (2023) The effects of pre-service early childhood teachers’ digital literacy and self-efficacy on their perception of AI education for young children. *Educ Inf Technol* 28(10):12969–12995

17. Shi H, Chen H, Yang Q et al (2023) A method for bio-sequence analysis algorithm development based on the PAR platform. *Big Data Min Anal* 6(1):11–20
18. Zhang K, Qu T, Zhang Y, Zhong RY, Huang GQ (2022) Big data-enabled intelligent synchronisation for the complex production logistics system under the opti-state control strategy. *Int J Prod Res* 60(13):4159–4175
19. Sumarliah E, Al-Hakeem B (2023) The effects of digital innovations and sustainable supply chain management on business competitive performance post-COVID-19. *Kybernetes* 52(7):2568–2596
20. Li Y, Wang Y, Wang L, Xie J (2022) Investigating the effects of stakeholder collaboration strategies on risk prevention performance in a digital innovation ecosystem. *Ind Manage Data Syst* 122(9):2045–2071

# Human-Centric Video Analysis in Industrial Environments



**Hayder Mohammedqasim, Roa'a Mohammedqasem, Bilal A. Ozturk, Habib Rahman Hamedy, and Ali bin Asghar**

**Abstract** Through creating AI systems that can interpret and grasp video content in a way that is similar to that of a human, human-centric deep video understanding seeks to close the gap between computer vision and human perception. Deep learning-based methods currently in use overlook the complexities of social relationships, human behavior, and emotional intelligence in favor of object detection, action recognition, and scene interpretation. Their inability to fully comprehend audiovisual content is hampered by this restriction. Human-centric deep video comprehension is an effort to bridge the gap between computer vision and human perception by developing AI systems that can comprehend and interpret video content in a manner akin to that of a human. When it comes to object detection, action recognition, and scene interpretation, deep learning-based techniques now in use ignore the intricacies of social interactions, human behavior, and emotional intelligence. This limitation makes it difficult for them to properly understand audiovisual content. Our method has applications in video analytics, social robotics, and human-computer interaction, among other areas. We can enhance AI systems' capacity to communicate with people, identify social signs, and offer more precise insights into human behavior by creating systems that can comprehend video information from a human perspective. The proposed method is Vision Transformers (ViTs) that is human-centric deep video detection. The method emphasizes on social interactions, human behavior, and emotional intelligence. The result achieve by methodology is 93% accuracy. (In order to develop AI systems that are more like humans, this abstract suggests a novel approach to deep video understanding that focuses on the subtleties of social interactions and human behavior.

**Keywords** People oriented · Video recognition · Artificial intelligence · Deep learning

---

H. Mohammedqasim · R. Mohammedqasem · B. A. Ozturk · H. R. Hamedy (✉) · A. Asghar  
Faculty of Engineering, Istanbul Aydin University, Istanbul, Turkey  
e-mail: [habibhamedy@stu.aydin.edu.tr](mailto:habibhamedy@stu.aydin.edu.tr)

## 1 Introduction

“Everyone is putting a big emphasis on artificial intelligence (AI) and it is all for a good reason: human-in-the-loop approaches are crucial for developing new video understanding technologies because they target the features most relevant to real-world scenarios [1].” Thus, understanding people and their interactions will help AI systems to decipher video content more effectively, thus enhancing technologies’ applicability in practical contexts. It is worth knowing that today machines have outsmarted humans in areas like face and object recognition, IQ tests, games, friendly speech recognition, written text comprehension, and translation, and much more [2]. The improvement of all these technologies is attributable to three basic pillars of technology. This entails access to ‘big data,’ such as thousands of hours of transcribed speech data and tens of millions of labeled image data. The second one is the nature of the availability of large computational capability like resources in terms of GPUs or clusters of cloud servers and so on [3]. Along with these two there are many other areas in ML which have observed many advancements like deep learning and reinforcement learning. This is probably why they could say that today we live in an era of [4]. Deep CNN solutions for still image classification tasks have been adopted for video classification tasks since 2012 [5]. However, recent advancements have introduced a new paradigm in AI: The nearest in this family is known as Vision Transformer or in short as ViT. Or rather while ViTs also employ self-attention to extract such long-term contexts as CNNs, the images themselves are processed in groups of patches rather than images. This has been applied and some promising results have been observed from various image classification ASLs including when it is used instead of standard CNN [6]. Also, Dosovitskiy et al. [6] states that Vision Transformers (ViTs) are providing much better performance in the recognition of image than other domains compared to traditional CNNs. Therefore, ViTs shows important improvements in different computer vision tasks.

### 1.1 Problem Statement

Even with tremendous advancement in AI, video understanding is far from being human-oriented. Due to the strict complexity that is inherent in human activities and interactions, major advancements in both AI and video understanding are needed. These traditional CNNs have a drawback of getting long-range dependencies and contextual information which are vital in video understanding. Self-attention-based Vision Transformers are equally efficient solution measures as well. To the best of knowledge of the authors, this concept is completely novel in the context of human-centered video analysis. This research is aimed at proposing a new approach that utilizes Vision Transformers to improve the performance of identifying and analyzing human activities and engagements in video to facilitate AI-based video analysis and security systems.

## 1.2 Contribution

The paper seeks to discuss Vision Transformers as a tool in enhancing human-centric video understanding through the use of deep learning techniques that have greatly enhanced AI's way of interpreting videos. Thus, the approach proposed in this paper is devoted to the notion of the gap between the computer vision methodology and the human perception of the world and its aspects, with the accent on the social interactions, people's actions, and feelings. The summarized key contributions of this research are the following:

- Utilization of Vision Transformers: Use of Vision Transformers on the video data. These transformers analyze the images and divide them into patches to carry out the self-attention for the right feature extraction and the analysis of the long-range dependencies within the video sequences.
- Human-Centric Video Analysis: Interested in the human aspects involved in action recognition, emotion detection, and social interaction analysis of the video, which provides more information about what the video is all about.
- Performance Metrics: Assesses the proposed approach using tradition metrics that is the accuracy, recall, precision, F1-score and AUC to compare the performance of ViTs in different video analysis tasks.
- Advanced Feature Extraction: Semi-supervised learning with few labeled samples and large number of annotated videos to enhance the accuracy and reliability of the obtained models for video analysis.
- Real-World Applications: Exemplifies the applicability of the developed method in the field of video analysis, social robotics, and HRI, which points to the potential for improving interaction between a human and a machine as well as the interpretation of people's behavior.

The work contributes a solid, dependable approach to human-centered video analysis, which points to the high growth rates obtained with Vision Transformers in the sphere of AI and video.

## 2 Related Works

Yu et al. [7] has presented a model that uses self-attention for sequence data. This helps the model to bring the word significance across the sentences and help the model identify the long dependency range which makes this technique very useful for sequence prediction tasks such as machine translation and sentiment analysis. ViT introduced the transformer architecture of Dosovitskiy et al. [6] for the image classification application. A ViT converts the images into patches and then applies self-attention to capture distant relations, and it also performs better than the conventional CNN models. Later, Fan et al. [8] generalized the ViT for object detection and semantic segmentation while performing better than the conventional CNN models.

In general, summing up at the global level, the strength of the approach was in the use of large annotated sets of data and in the integration of the self-attention mechanism of ViT which in its turn had a positive impact on the increase of the accuracy of objects detection and the outlines segmentation. In another related work, Landon et al. [9] integrated attention mechanisms into ViTs with the goal of discovering social interaction and emotions. Thus, integration of end-to-end with multiple levels of social signal processing gives a better understanding of the social dynamics and offers a more accurate interpretation of the social interaction in the videos. In a more recent work, ViTs were used to enhance the MPT and HPE in complex environments. Particularly, when using ViT combined with the basic tracking algorithms but with different treatment approaches such as illumination, motion variation, tracking effectiveness, and light change robustness there is an improvement Bao et al.[10]. In this regard, Lubana et al. [11]expanded the layer normalization to deep learning models and, more especially, to ViTs. Layer normalization was proposed to normalize the activations at the output of each layer during training to keep the mean and variance fixed. Layer normalization was able to help in stabilizing the deep learning models and also increase the rate of convergence. Bayoudh et al. [12] associated computer vision and deep learning to the capability and effectiveness of ViTs with regard to multimodal data, which means text and image data. Therefore, ViTs, as well as other transformer models, are effective and rather appropriate for various types of data and, therefore, enhance the overall performance. Vesnin et al. [13] in their research work explained how ViTs can be used in semantic similarity search while comparing the documents and patents. The information obtained from the Vision Transformers is capable of extracting multiple semantic aspects of the provided sentences, which in turn enables providing more precise results of the semantic similarity in contrast to the approaches used before. Taken together, these works collectively imply, on one side, that the Vision Transformers have developed through and are progressively being supplanted by understanding video and the object detection and, on the other side, by the development of human-oriented AI systems in a manner that they offer the capacity to help improve the precision as well as efficiency of AI models on a broad range of complex tasks. In addition, some related works have also demonstrated that ViTs are useful in the tasks of autonomous driving and medical image segmentation. ViTs such as the BEVFormer and CrossDTR show improved performance in 3D object detection and tracking because of the self-attention mechanism for capturing spatial and temporal dependencies Lai-Dang et al. [14]. Likewise, the hybrid model TransUNet which combines the CNNs and the ViTs performs excellently in medical image segmentation through the local feature extraction from the CNNs and the global context from the Transformers. This model performs the best with the Dice Similarity Coefficient benchmark dataset, which indicates its capability to segment complex medical images and efficiency on various modalities Islam et al. [15]. Such developments show that with the help of ViTs it is possible to enhance the spatial orientation and the efficiency of the model in different tasks.

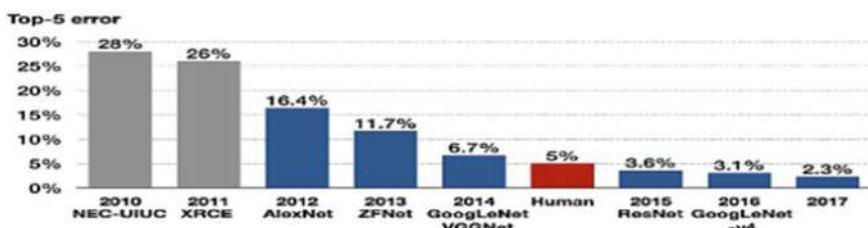
### 3 Methodology

Generalizations such as the one in the heading are that normally, if there is a great leap in one of the two broad technological areas of still images, there will not be much time before similar improvements are made also in the field of moving pictures. Video analytics to support intelligent applications are well-accepted in the enterprise as well as the consumer sectors. Unlike the more obvious area of public security, new, and upcoming uses include the business sector, home security, self-driving cars, and narration.

They also presented a relatively large number of difficulties in comparison with images in general. It is even more challenging when one is detecting pedestrians in videos rather than in still images as videos encompass almost all the possible content that could be conceived. Storage, computation, and communication are the major requirements that are greatly demanded in today's world. At times there may be a necessity to process the information as it comes in or is provided. The general labeling of the video data is specifically expensive while in some occasions, we do not have sufficient training data. For instance, in the case of desire to analyze surveillance video data, access to the data is frequently problematic and the number of positive samples may be low. Such challenges push the availability of video analytics technologies into practical application to the limits. On the other hand, improvement in deep learning methodologies and designs has decreased the error rate over last decade. Figure 1 shows the decrease in Top-5 error rate in image classification that improved model performance.

Thus, more sophisticated approaches appear to be used to improve video understanding technologies due to the efforts needed to reduce their complexity and the variations within the video material. In this context, the Vision Transformer (ViT) has found to be quite efficient in terms of this concern. The models that incorporate ViTs with self-attention mechanisms are also higher in power than the CNN and LSTM models to capture the nature of both spatial and temporal features from the given datasets of videos.

Vision Transformers can accept sequences of image patches as input, therefore, Vision Transformers are useful for video processing when temporal dynamics analysis is considered. They are able to handle big datasets and large amounts of computations hence fulfilling the big data and computational requirements. Due to their



**Fig. 1** Performance of the winners of the ImageNet classification competitions over the years

ability to process both spatial and temporal characteristics, they can be used in real-time object detection and decision making for self-driving cars have shown that these systems can efficiently combine the advantages of ViTs for improving the autonomous driving technologies [16]. Furthermore, because the ViTs are good at the feature extraction step, they can be fine-tuned with the few labeled samples as a result of the limitation of labeling cost and the availability of small training data.

This novel approach based on Vision Transformers can immensely contribute to the future advancements in video understanding due to the significance of the following fields: real-time video analysis, self-driving vehicles, and multi-tier security cameras. Thus, the focus on implementing human-centric features complemented with the strengths of ViTs can help address many of the fundamental issues relating to video understanding, opening up avenues for the development of sophisticated and practical video analytics [17].

#### Human-Centric: Why and How?

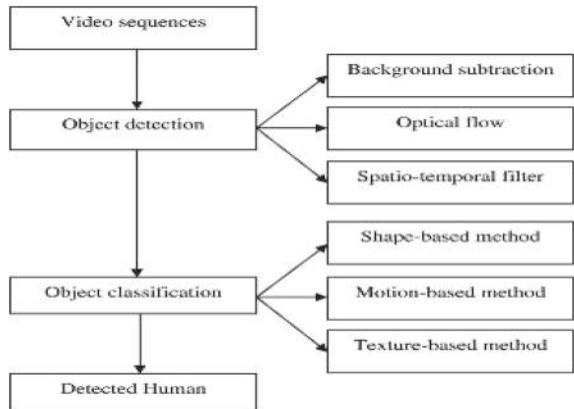
Individuals really form the core, soul, or essence of our everyday living and our organizational practice. To serve people better, we need to understand them more deeply: them and their environment, and information about their identity, activities, actions, emotions, plans and purpose of a conversation, and interlocutor. As people are always searching for themselves and searching for other people, machines also require effective guidance modes to enable them to perceive people through multi-sensory information as the intelligence of society continues to rise.

As expected, people appear in most of the contents of the videos, a key area of human hybrid intelligent computing is human–computer interactions which has been almost continuously likely to remain the most dominant in practical applications in the near term. That is why human intervention takes an important position while analyzing the given video and interpreting its content. Recognizer is possibly one of the first successful applications of computer vision where it tackled face identification or recognition. As far as the next discovery is concerned, it is perhaps not impossible that it may emerge from the general human understanding technologies. Taking all into account, further video comprehension it is reasonable to go with the human-oriented one as the result. When we talk about the ‘human-centric’ aspect, we are going to include the tasks of human detection in video and design principles based on the constitution of brain circuitry, albeit the current state is not so advanced [14].

The original intention of AI was to do exactly what this model was intended to do, recreate the functioning of the human brain. It is important to understand the human from the various facet and fields such as Biology, Neuroscience, cognitive sciences, behavioral sciences, and social sciences. In this context, one perhaps the most promising technique for video understanding is deep learning. Among the groups, Vision Transformers (ViTs) have now come up as a tool of utilizing self-attention mechanisms that mimic the brain of a human. This mechanism has been integrated into some of the design of neural networks that allow a model to pay more attention to some features or the relationships with data.

Specifically, ViTs are quite impressive to capture spatial and temporal patterns, which are significantly helpful and useful in human-orientated activities such as

**Fig. 2** Flow chart for the proposed model



action recognition, emotion detection, and human interaction analysis. Therefore, through the employment of attention models that mimic human cognitive paradigm, Vision Transformers have presented themselves as fitting models to analyze detailed data from videos.

Further, the argument of human beings is grounded on the fact that one is bound to acquire knowledge in the process of reasoning more than anything else. Nowadays, it has been realized that there has been more enhancement of knowledge in data savvy methodologies in deep learning system designs. This integration makes it easier to enhance the capacity of the AI systems in analyzing individuals actions and actions in frames of videos well.

Next section, describes the human-centric vision tasks and explored the latest status of them, the challenges and the current limitations of how the authors' imperfect understanding of the human brain working mechanisms, such as attention mechanisms, semantic models, and knowledge-based reasoning, can be applied to address some of these challenges more effectively.

In Fig. 2, the flow chart explains the general procedure for object detection and classification in order to recognize people in video streams. At first, there is the input video sequences, then the object detection methods, and finally, the classification methods are to identify whether the detected objects are humans or not. It is useful for accurate identification of humans in video data through integrating several detection and classification procedures [18–24].

## 4 Results

Object recognition in a video depends on the identification of people and their characteristics and what they are doing in the scene. Much has been achieved in many fundamental vision tasks with the focus on human-centric applications. The below are shown result using Tables and figure.

Table 1 shows the different techniques and parameters used in phases of model training and achieved accuracy.

Table 2 shows CV Fold that is the model performance on different subsets.

Table 3 shows performance of ViT Model, in term of correctness of accuracy, precision, recall, F1-score and Receiver Operating Characteristic Area Under the Curve (ROC AUC).

**Table 1** Result and parameters for the proposed model

Technique	Phase-1	Phase-2
Fine tuning	Learning rate: 0.001	Epochs: 5
Data augmentation	Random rotation: 0–30	Flipping: Yes
Advanced preprocessing	Normalization	Mean:0.5, Std Dev: 0.2
Ensemble method	Decision tree	Random forest: 10
Cross validation	Folding	CV Fold:5
Final accuracy	92.5%	93%

**Table 2** Accuracy by using different CV Fold score

Metric	Value
CV score	Fold 1
CV score	Fold 2
CV score	Fold 3
CV score	Fold 4
CV score	Fold 5
CV score	Fold 6
CV score	Fold 7
CV score	Fold 8
CV score	Fold 9
CV score	Fold 10
Mean CV accuracy	0.933
Test accuracy	1.0

**Table 3** Performance metrics the of ViT model

Evolution metrics	Percentage
Accuracy	86.66
Precision	89.65
Recall	83.87
F1 score	86.66
ROC AUC	92.66

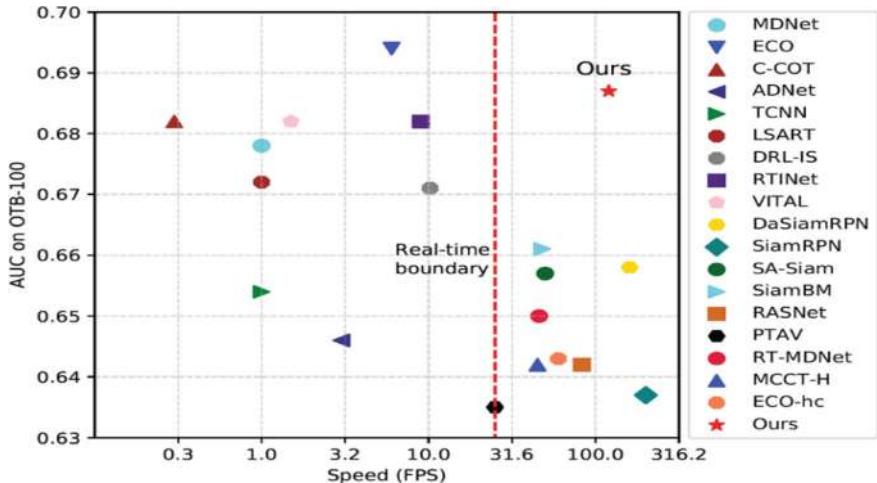
## 4.1 People Tracking

Computation of Visual Object Tracking is one of the universally important issues discussed under video analysis and understanding. It is the job of a tracker to locate the target object within any of the frames of a video once the bounding box of the target object has been provided in the first frame of the video. We can generally regard single object tracking as a detection tracking problem because in essence it involves tracking detection of this particular object from one frame to the other. The major challenge lies in the fact that at the same time it is necessary to respond to two opposite although partly contradicting demands—on the one hand the focus on the ability to identify unfamiliar classes on the other hand the ability to distinguish between the classes already seen. Here robustness comes into the picture when a tracker should not lose track of the target even if there is variation in illumination the target is in motion or change in view angle or object deformation. On the other hand in the case of a tracker it has to segregate the target object from another similar object plus or a number of objects in the background, etc. Both are conventionally managed through online training that strives to attain flexibility.

Since 2015 there has been some research where authors employ CNNs in their work. Although deep features are disadvantageous in the sense that they slow down the speed of online training in some way this very advantage of deep features allows one to completely eliminate online training. The first work in this regard is SiamFC where the Siamese CNNs are utilized to extract the features of the target and the search region and followed by conventional cross-correlation layer which provides an efficient sliding window mechanism. Therefore SiamFC achieves the working frame rate of 86fps on the GPU side. Though the basic concept can be attributed to this paper there are many follow-up works like SA-Siam and SiamRPN ([Fig. 3](#)).

Most of the object tracking work which utilize SiamFC has a single-stage architecture while a two-stage SiamFC-based network has been introduced to solve for robustness and discriminativeness. Above all, this paper investigates two stages: the first coarse matching stage aims to improve robustness and the second fine matching stage aims to increase the discrimination power by replacing traditional cross-correlation layer with more sophisticated distance learning subnetwork. These results are statistically sound and the tracker operates at a phenomenal 120 framerates based on an NVIDIA P100 GPU. Some of these comparisons have been highlighted in [Fig. 3](#) which illustrates the statics of the benchmark OTB-100 to show the best trackers with high accuracy and low speed.

In various real-life situations, the need appears to track several individuals at once. In general, such an approach is not very efficient, when multiple-person tracking is solved as a sum of multiple single-person tracking problems. Modern multi-person tracking approaches rely on the tracking-by-detection paradigm, which means using a general object/person detector to detect objects (of the same target class, which is person) in distinct frames and then connecting the detections. These include; importance sampling and particle filtering for the propagation of the state in a Bayesian way, linking the short tracks over long time using the Hungarian Algorithm for the



**Fig. 3** Accuracy-speed trade-off of top-performing trackers on the OTB-100 benchmark. The speed axis is logarithmic

optimal assignment and the greedy Dynamic Programming in which the trajectories are put one step at a time. Recent studies to increase the reliability of wrong identity assignment have posed the task of connecting detections for a longer period using optimization techniques. A typical method in multi-person tracking problem is the use of constrained flow optimization they use the k-shortest paths method. Other related approaches include graph-based formulations of the minimum cost multicut problem.

The work compared the tracking-by-detection approach that separates tracking from detection to a tracking-by-detection approach even though the former may not be the most efficient method, a lot more research should be dedicated to joint detection and tracking approaches applied to multiple-object/people tracking. Multiple-Object/Pet Multiple People Interactions: More use of spatial and temporal structures needs to be made; the balance between the complexity and accuracy has to be found. For long-term tracking, it becomes even more complex and often provides intermediate short-term tracklets, and completing linking/matching tracklets, over time, it becomes necessary, for instance, by using object re-identification (re-ID) methods.

## 4.2 Human Pose Estimation

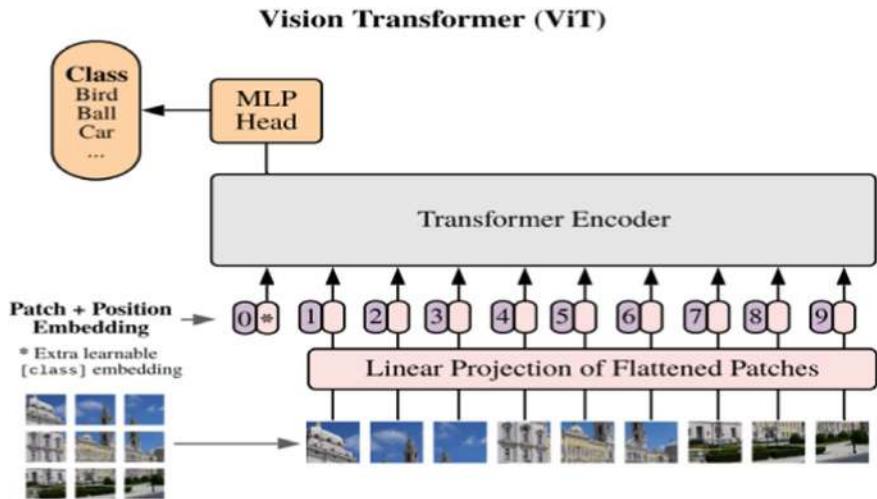
Human pose estimation deals with determining the position of features in an image which are corresponding to specific body joints in a human figure. Human body is of course an important application; and it has large number of applications in many fields including human action recognition motion analysis activity analysis and

human computer interaction. Nonetheless progress there is still a very challenging task and despite the fact that many approaches have been already proposed and some techniques have got particularly much attention during the last several years there are a number of factors which challenge anybody who tries to develop a successful pose estimator: poses shapes views sizes dependencies between those parts appearances and qualities of images.

The pictorial structure is one of the earlier works that instantiate deformable configurations as all the parts are connected to each other by springs to model complex joint relationships between the pairs of parts. As for subsequent works this concept is generalized to CNNs. Most of the recent methods in this type of framework train CNNs to learn feature representations as well as to get the positions of the 2D joints or the score maps of the 2D joints. Some methods combine the learned feature representations to directly regress the 2D joint positions. Some methods estimate a score map for each 2D joint based on architectures fully CNN is a well-mapped way. For effective sampling of the multiple people poses in 2D some methods use Part Affinity Fields to train to link body parts to people in the image. The architecture incorporates the global context aggregation making it possible to implement greedy bottom-up process of parsing while using real-time speed and without losing its high accuracy even when the number of people shown in the picture is rather great. Body normalization first and then limb normalization reduces the movement variance of the relative joint locations which benefits the learnability of convolution spatial models and improves the pose estimation accuracy on the downside it is computed in the 3D space leading to higher memory usage.

### 4.3 Person Re-Identification

Person re-identification (re-ID) using Vision Transformers (ViT) involves finding a specific person from the multiple camera angles or times, or from the same angle but at different time. This task is considered difficult due to changes in person pose, viewpoint, detection, background, occlusion, and lighting. Subsequent work has been devoted to addressing these problems by building upon what ViT excels in addressing, which are spatial misalignment and semantic alignment. One proposed solution is to use ViT's architecture to split the person image into non-overlapping patches which allows to learn densely semantically aligned features. This makes it possible for semantically aligned model construction and analysis of semantically aligned feature learning. To tackle with the issues like errors and the difficulties in handling the non-overlapping areas a learning paradigm can be introduced to learn the features with the help of another stream [9]. Otherwise, an encoder-decoder structure can be applied: On the contrary, the encoder employs ViT to acquire the re-ID features of the input image, and the decoder synthesizes a 3D full-body texture image in the canonical semantic space. This ensures that the learned features are invariant to view and pose, thus explaining the elimination of Fig. 4 visible body discrepancies in matching images. The decoder is only employed during training while it does not



**Fig. 4** Vision transformer process explanation

contribute to the model's complexity during the testing phase. Integrating person re-ID in ViT holds promising as it can take advantage of the approach and detect space and hierarchical components besides the semantic alignment.

## 5 Conclusion

To sum up, for a better and comprehensive understanding of videos and video sequences, human-oriented vision tasks must be incorporated. This is the area in which the system perspective is most helpful, as one can effectively combine advantages in one BB and constructively address possible risks in others. Depending on the application, a practical system may integrate some or all of the following: Some applications include person detection and tracking, re-identification, pose estimation, action recognition, detection of heatmaps, etc. For instance, in a retail intelligence with several cameras as depicted in Fig. 3, there may be the need to track or recognize a customer through face detection, body only detection, bones only or both. The longitudinal linkages are important in joining tracklets of the similar time period and the same person across different cameras through cohort analysis and person re-ID, respectively. This way, heatmaps might be created while tracking the people and to know more about the customer. It's at this point where we can use the estimated pose sequences or even pose and RGB data to identify more of the detailed activities. Special consideration has to be paid to dependency constraints when dealing with real-time interactive application like video teleconferencing is very demanding in terms of size can be as low as 100k bytes and also very strict on speed in terms of ms per frame. Therefore, the task should be shifted to the reduction of model size, and

the methods that are focused on accelerating the ViT-based models include knowledge distillation, model pruning, and quantization. Hence, incorporating ViT with other human-centric vision tasks will be able to give a complete outlook of the video comprehension and yield a more topnotch system level result. Future work, further investigation is required to combine ViTs with XAI for safety-critical tasks as well as to design efficient training techniques to address computational challenges which would allow the usage of ViTs in low-resource environments.

## References

1. Zeng W (2020) Toward human-centric deep video understanding. *APSIPA Trans Signal Inf Process* 9:e1. <https://doi.org/10.1017/AT SIP.2019.26>
2. Chen W et al (2023) From gap to synergy: enhancing contextual understanding through human-machine collaboration in personalized systems. In: UIST 2023—proceedings of the 36th Annual ACM symposium on user interface software and technology, Oct 2023. <https://doi.org/10.1145/3586183.3606741>
3. Braun A, Warth G, Bachofer F, Schultz M, Hochschild V (2023) mapping urban structure types based on remote sensing data—a universal and adaptable framework for spatial analyses of cities. *Land* 12(10):1885. <https://doi.org/10.3390/LAND12101885>
4. Paletta L, Paar G (2002) Bayesian decision fusion for dynamic multi-cue object detection
5. Wang L et al (2016) Temporal segment networks: towards good practices for deep action recognition. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), vol 9912, pp 20–36. [https://doi.org/10.1007/978-3-319-46484-8\\_2/FIGURES/3](https://doi.org/10.1007/978-3-319-46484-8_2/FIGURES/3)
6. Dosovitskiy A et al (2020) An image is worth 16 x 16 words: transformers for image recognition at scale. In: ICLR 2021—9th international conference on learning representations, Oct 2020 [Online]. Available: <https://arxiv.org/abs/2010.11929v2>. Accessed 26 May 2024
7. Yu Z, Peng S, Niemeyer M, Sattler T, Geiger A (2024) MonoSDF: exploring monocular geometric cues for neural implicit surface reconstruction. [Online]. Available: <https://nijinshuchong.github.io/monosdf>. Accessed 26 May 2024
8. Fan Z, Hu G, Sun X, Wang G, Dong J, Su C (2022) Self-attention neural architecture search for semantic image segmentation. *Knowl Based Syst* 239:107968. <https://doi.org/10.1016/J.KNO SYS.2021.107968>
9. Landon LB et al (2019) The behavioral biology of teams: multidisciplinary contributions to social dynamics in isolated, confined, and extreme environments. *Front Psychol* 10:442358. <https://doi.org/10.3389/FPSYG.2019.02571>
10. Bao Q, Liu W, Cheng Y, Zhou B, Mei T (2021) Pose-guided tracking-by-detection: robust multi-person pose tracking. *IEEE Trans Multimed* 23:161–175. <https://doi.org/10.1109/TMM.2020.2980194>
11. Singh Lubana E, Dick RP, Tanaka H (2021) Beyond batchnorm: towards a unified understanding of normalization in deep learning. *Adv Neural Inf Process Syst* 34:4778–4791
12. Bayoudh K et al (2021) A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. *Vis Comput* 38(8):2939–2970. <https://doi.org/10.1007/S00371-021-02166-7>.
13. Vesnin D, Levshun D, Chechulin A (2023) Trademark similarity evaluation using a combination of ViT and local Features. *Information* 2023 14(7):398. <https://doi.org/10.3390/INFO14070398>
14. Lai-Dang Q-V (2024) A survey of vision transformers in autonomous driving: current trends and future directions, Mar 2024. [Online]. Available: <https://arxiv.org/abs/2403.07542v1>. Accessed 26 May 2024

15. Islam K (2022) Recent advances in vision transformer: a survey and outlook of recent work, Mar 2022. [Online]. Available: <https://arxiv.org/abs/2203.01536v5>. Accessed 26 May 2024
16. Li Y et al (2023) VoxFormer: sparse voxel transformer for camera-based 3D semantic scene completion. [Online]. Available: <https://github.com/NVlabs/VoxFormer>. Accessed 26 May 2024
17. Nicoliciu AM, Nicoliciu AL, Alexe B, Teney D (2023) Learning diverse features in vision transformers for improved generalization, Aug 2023. [Online]. Available: <https://arxiv.org/abs/2308.16274v1>. Accessed 26 May 2024
18. Mohammedqasem R, Mohammedqasim H, Ata O (2022) Real-time data of COVID-19 detection with IoT sensor tracking using artificial neural network. Comput Electr Eng 100:107971. <https://doi.org/10.1016/J.COMPELECENG.2022.107971>
19. Mohammedqasim H, Ahmed Jasim A, Mohammedqasem A, Ata O (2024) Enhancing predictive performance in COVID-19 healthcare datasets: a case study based on hyper Adasyn over-sampling and genetic feature Selection. J Eng Sci Technol 19(2):598–617
20. Alyasin EI, Ata O, Mohammedqasim H, Mohammedqasem R (2024) Enhancing self-care prediction in children with impairments: a novel framework for addressing imbalance and high dimensionality. Appl Sci 14(1):356. <https://doi.org/10.3390/APP14010356>
21. Mohammedqasim H, Mohammedqasem R, Ata O, Alyasin EI (2022) Diagnosing coronary artery disease on the basis of hard ensemble voting optimization. Medicina 58(12):1745. <https://doi.org/10.3390/MEDICINA58121745>
22. Jasim AA, Hazim LR, Mohammedqasim H, Mohammedqasem R, Ata O, Salman OH (2024) e-Diagnostic system for diabetes disease prediction on an IoMT environment-based hyper AdaBoost machine learning model. J Supercomput 1–26. <https://doi.org/10.1007/S11227-024-06082-0/TABLES/4>
23. Mohammedqasem R et al (2023) Multi-objective deep learning framework for COVID-19 dataset problems. J King Saud Univ Sci 35(3):102527. <https://doi.org/10.1016/J.JKSUS.2022.102527>
24. Alyasin EI, Ata O, Mohammedqasim H (2022) Novel hybrid classification model for multi-class imbalanced lithology dataset. Optik (Stuttg) 270:170047. <https://doi.org/10.1016/J.IJLEO.2022.170047>

# Survey and Analysis of Communication and Routing Mechanisms for UAVs in Wireless Sensor Networks



Prajoy Podder  and Maciej Zawodniok 

**Abstract** Unmanned aerial vehicles (UAVs), commonly called drones, have seen much relevance in most disciplines because of their multiple uses such as in military operations, spying, monitoring the impact of climate change, and delivery. Emerging UAV communication technologies addressing the applications in next-generation wireless networks have been explored in this paper. Specifically, it investigates networking technologies that communicate successfully between UAVs, and routing strategies in UAV-assisted wireless sensor networks (WSNs). Networking and common communication technologies for UAV communication systems have been discussed in this survey paper. Drawing insights from recent academic and industrial literature, we have provided a comprehensive overview of UAV communication technologies, including IoT-enabled communication, ultra-reliable low-latency communication (URLLC), navigation strategies, and advancements driven by machine learning and artificial intelligence.

**Keywords** Unmanned aerial vehicle (UAV) · Flying ad-hoc network (FANET) · Wireless networks · Multi-UAV system · Latency · Communication systems · Routing protocols · Energy efficiency

## 1 Introduction

UAVs are gaining popularity across numerous industries, leading to an expansion of their accessibility in consumer markets. They are commonly utilized in various sectors, including security, agriculture, forestry, surveillance, and transportation, for critical missions such as rescue and environmental conservation.

---

P. Podder () · M. Zawodniok

Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO, USA

e-mail: [pp64k@mst.edu](mailto:pp64k@mst.edu)

M. Zawodniok

e-mail: [mzawodniok@mst.edu](mailto:mzawodniok@mst.edu)

Drone communications employ a variety of network protocols and wireless channels, which are determined by the application. In outdoor environments, point-to-point communication via direct line-of-sight is frequently adequate to ensure uninterrupted signal transmission. On the contrary, satellite communication connections are frequently employed by drones for the purpose of surveillance, specifically in operations related to security, defense, or extensive outreach. Preference is given to cellular communication technologies for civil and personal applications. For the sake of efficacy, indoor communication, especially in mesh networks and wireless sensor networks (WSNs), typically employs point-to-point protocols like Bluetooth. The complexity of the communication process increases when it is implemented in multi-layered networks. The utmost importance of data transmission requires the implementation of suitable routing protocols. Networking is an essential characteristic for drone clusters and individual UAVs [1–4]. The integration of drones into a multitude of communication systems, such as vehicular communication systems, wireless sensor networks, and mobile communication networks, has significantly broadened their range of applications in conjunction with the Internet of Things (IoT) [5].

Several researchers have integrated artificial intelligence, cryptography, and navigation strategies into unmanned aerial vehicle (UAV) communication protocols in order to guarantee dependable, efficient, and instantaneous correspondence among nodes in the UAV network [6, 7]. However, energy efficiency and drone speed must take precedence to guarantee dependable and secure communication. Challenges expressed by drones may stem from problems that originate from the low power and computing capabilities of a system [8]. Academics have therefore given insightful analysis on how to achieve better resolutions for such issues. Also, the interruptions of communication through interference of the aerial network are another issue of concern. The interference has the propensity of overpowering UAV networks which are gradually being deployed for aerial surveillance and communication disaster response infrastructure as described in [9, 10].

## 1.1 *Questions to Address*

The scientific question we aim to address in this paper is: “What are the emerging UAV communication technologies, their advantages, use case scenarios, technical challenges, and future directions, and how can they be applied to enhance next-generation wireless networks?”.

## 1.2 *Research Motivations*

Ensuring efficient communication for coordination and cooperation is the major difficulty in multi-UAV system design. UAVs can be used in aerial sensor networks, in

which nodes gather information from several sources spread across different places. It can include different types of sensors, each of which needs a different form of data transmission system. However, there are problems with UAV networks as well: low bandwidth, mobility, erratic connectivity, short transmission range, and unanticipated noise [1, 11, 12]. Upholding transmission range between two UAVs moving at high speeds in opposite directions could be difficult.

### 1.3 Contributions

In this review paper, we perform a survey of the multiple UAV communication technologies to investigate their advantages, the fields of application, technical challenges, and further developments. This survey covers the communication and network technologies specific to UAVs that include research on appropriate task modules, antennas, resource management, and network architectures.

## 2 Literature Reviews

Technical survey report in [13] deals with the UAV traffic monitoring and management which also includes the pros and cons of such systems. In [14], the classification of UAV network was proposed based on formation where single UAV and multiple UAV were considered, and several applications discussed for both classifications were also presented. Authors classified multi-UAV systems according to the distribution parameters such as cooperative networks and self-organizing networks and motivated the readers with software solutions, simulators, and open issues.

The requirements and characteristics of UAV networks appropriate for civilian use were described in [11] by categorizing UAV networks into communication-based application areas of search and rescue and delivery of goods. They also described the issues and constraints concerning QoS and quantitative communications and assessed the potential of the existing wireless technologies for UAV networks. In [12], He et al. focused on the concern of security factors relating to UAV's communications, on condition of GPS and WiFi, where some attacks like Jamming and Spoofing were applied and new defense measures were suggested. In [15], fundamental networking topologies were reviewed; especially the channels UAV-ground and UAV-UAV, as well as innovation on UAV communications where discussed together with proposals on how the performance can be improved. In [16], authors have categorized the functions and services required to support the UAV-based system, functions, and services of architecture requirements, and the middleware layer services of communication which supports communication between different interfaces of UAV-based systems. They also discussed about the UAV usage in data gathering from WSN. In [17], the

authors have described the heads of UAV networks in terms of wireless communication and presented communication categories, technologies, assessment methodologies, and open issues. The analysis of FANET architectures along with routing techniques and the proposed taxonomy for the routing protocols were discussed in [18], where major open issues were also presented.

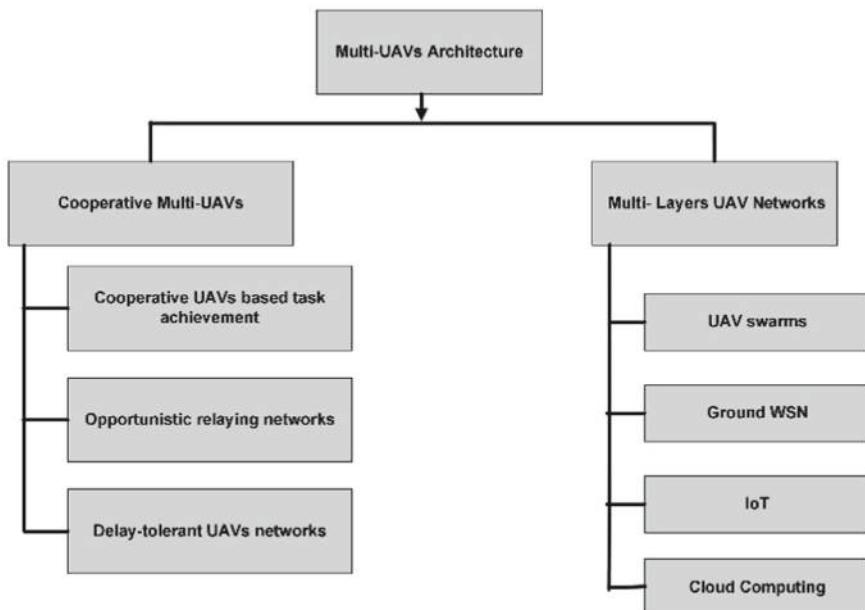
In [19], authors provided information about the background of FANET, communicated challenges, mobility models, trajectory optimization, and security aspects. In [20], the wireless network classifications concerning the UAV network characteristics, the design problems, the routing protocols, the QoS-based routing protocols, and the security in terms of QoS metrics have been presented. Comparatively, in [21], the authors focused on security and privacy concerns to create FANETs using UAVs. They [21] also talked about threats that exist to the Internet of Drones (IoD) which is some sort of network framework for managing and granting access over the Internet between UAVs and users.

### 3 UAV Architectures

UAVs can be categorized into single and multi-UAV systems. Single UAV systems have shown remarkable adaptability and efficacy in a variety of applications, including area surveillance [12–14, 17], target tracking [13, 14], transportation [11, 13], agricultural, search and rescue [11, 17, 19, 20] and others. Because of inherent limitations in its design and operating capabilities, using a single UAV in many applications has significant drawbacks. For instance, the efficiency of activities over huge agricultural fields is reduced because a single UAV can only cover a limited area during each flight. UAVs must have great stability and accuracy for jobs like infrastructure inspections to guarantee correct data collecting. The stability of a single UAV may be compromised, particularly in inclement weather, which would reduce the precision of the data obtained. On the other hand, multi-UAV systems can cover larger areas in less time, making them indispensable for large-scale operations such as agricultural monitoring of vast farms or rapid assessment of post-natural disasters. A single UAV can be considered inefficient sometimes in critical applications such as search and rescue (SAR) or military operations. Multi-UAV systems provide redundancy, enhancing reliability and operational safety by ensuring that the failure of one UAV does not cripple the mission. Table 1 summarizes the performance comparisons of single and multi-UAV systems [22]. Multi-UAV architecture can be divided into groups depicted in Fig. 1.

**Table 1** Performance comparisons of a single UAV and multi-UAV system

Features	Single UAV [22]	Multi-UAV [22]
Survivability [22]	Poor	High
Autonomy [22]	Low	High
Speed of mission [22]	Slow	Fast
Scalability [22]	Limited	High
Cost [22]	High	Low
Communication needs [22]	High	Low
Radar cross sections [22]	Large	Small

**Fig. 1** Multi-UAV architectures

## 4 Communication and Network Technologies for UAVs

The integration of 5G technology has emerged as a prominent trend in UAV communication, offering higher data rates, lower latency, and increased reliability. This enables UAVs to transmit and receive large volumes of data more efficiently, opening up new possibilities for real-time applications. This article discusses how multi-UAVs may be used for various operations and purposes.

## 4.1 Integration of UAVs in Wireless Sensor Networks and Vehicular Communication Systems (VCS)

Incorporating drones into WSNs poses challenges due to dense sensor placement across large areas. Table 2 describes the literature survey in which UAVs are integrated into WSN and VCS.

## 4.2 Integration of UAVs in Cellular Networks

The integration of unmanned aerial vehicles (UAVs) into cellular networks has evolved significantly from early experiments with GPRS in 2006 [30] to advanced LTE-UAV networks tested by major organizations like China Mobile and Ericsson by 2016. These developments have led to various field studies, particularly those conducted under the auspices of the 3rd Generation Partnership Project (3GPP), which concluded in 2017 with detailed technical reports. These studies primarily focused on understanding how UAVs interact with existing cellular networks, revealing several challenges [31–34]:

- (a) *High Line of Sight Interference*: UAVs experience more interference than terrestrial users due to their elevated positions, which expose them to signals from multiple cell sites.
- (b) *High Altitude Issues*: UAVs often receive signals from the side lobes of downward-tilted Base Transceiver Station (BTS) antennas, potentially connecting to more distant BTSs rather than the nearest ones.
- (c) *Measurement Reporting Differences*: The metrics for signal strength and quality (RSRP and RSRQ) differ significantly between UAVs and ground users due to the UAVs' airborne nature.
- (d) *High Mobility*: UAVs' frequent movement leads to more signal handovers and less stable connections compared to terrestrial users.

## 5 UAV Routing Protocols

UAV networks, a subcategory of Mobile Ad Hoc Networks (MANETs), face unique challenges due to highly mobile nodes, fluctuating link quality, and limited energy on each UAV. To address these issues, researchers have adapted existing MANET routing protocols and developed UAV-specific protocols. These protocols fall into three main categories: topology-based, position-based, and swarm-based, as described in Table 3.

**Table 2** Survey of UAV-assisted wireless sensor networks and vehicular communication systems

Ref	Key contributions	Limitations
[23]	Explored diminishing effectiveness of static WSNs during disaster progression and proposed recommendations based on disaster management stages	Did not extensively explore the detailed technical aspects of UAV technology or the complexities of wireless communication protocols in disaster scenarios
[24]	Suggested UAV-based mobile data gathering in WSNs, with a proposed routing scheme for Route Selection and Communication Association (RSCA) using a regulated greedy algorithm	The proposed algorithm has a maximum deviation of 154% and an average deviation of 145% from the optimal solution, indicating room for improvement in achieving closer optimality
[25]	Presented and reviewed D2D advancements which can be an efficient approach for inter-UAV communication	Need for further exploration of the potential security vulnerabilities associated with D2D communication in critical scenarios
[26]	Proposed UAV-assisted vehicular networks (DAVN) which can enhance infrastructure coverage, vehicle-to-vehicle connectivity, network inter-working efficiency, and data collection	The absence of practical solutions in DAVN integration highlights the need for future efforts to efficiently integrate drones into vehicular networks. Exploring Software-Defined DAVN, which decouples the control plane and data plane, can facilitate network reconfiguration and address this challenge effectively
[27]	Introduced UAVR-S and UAVR-G protocols for UAV-assisted VANETs. Adopted the IEEE 802.11p MAC protocol for both inter-vehicular and UAV-to-Vehicle communication	Assumed that UAVs maintain a low constant altitude that does not exceed 200 m during the flight. Considered urban environment
[28]	Presented Vehicle-Drone hybrid vehicular ad-hoc Network (VDNet), utilizes UAVs for boosting data transmission between vehicles and achieves significant performance	Security challenges in infrastructure-less UAV-assisted sparse VANET systems include unauthorized access, data integrity risks, privacy concerns, traffic analysis, and physical vulnerabilities, requiring robust measures for protection
[29]	Suggested a smart drone for FirstNet application that relied on multi-hop D2D communication. Simulation results help in proving the paradigm necessity of UAVs depending on the distance and the transmit power values	Did not address the potential drawbacks such as latency, and scalability associated with implementing a drone-assisted multi-hop device-to-device communication scheme for extending network coverage

**Table 3** Advantages and limitations of various UAV routing protocols

Category	Sub-category	Example of protocols	Advantages	Limitations
Topology based	Static	DCR [35]	Minimizes control message exchange, and can adapt to changing network conditions, including the addition or removal of UAVs	Temporary delays in data delivery, limited mobility
		MLHR [36]	Reduces the routing burden on individual UAVs by arranging the network into clusters, with cluster heads managing communication within clusters and between clusters	Single Point of Failure, Overhead for Cluster Management
		LCAD [37]	Improved security and optimized delivery routes by minimizing travel time and energy expenditure	Limited Availability of Information, computationally expensive
	Proactive	OLSR [38]	Scalability, reduced overhead, fast convergence, multi-hop routing	Selecting and maintaining MPRs can introduce some additional overhead and security threats
		DSDV [39]	Scalable, efficient route updates, good for stable networks	Slow convergence, and high routing table overhead, are not ideal for highly dynamic networks
	Reactive	DSR [40]	Loop-free routes, adaptable to network changes, good for unpredictable mobility	High overhead for route discovery, complex route maintenance, and scalability concerns in large networks
		AODV [41]	On-demand route discovery reduces overhead in stable periods,	Route discovery overhead during route breaks, potential for routing loops, less efficient for frequent topology changes

(continued)

**Table 3** (continued)

Category	Sub-category	Example of protocols	Advantages	Limitations
	Hybrid	ZRP [42]	Reduced Overhead, Efficient Route Maintenance using a zone mechanism	The limited size of the zone is insufficient for accommodating high levels of mobility
		TORA [43]	Allows for the discovery and maintenance of multiple routes between source and destination nodes. This provides redundancy and improves fault tolerance	TORA relies on synchronized clocks among nodes in the network which often considers challenging to achieve in dynamic environments like UAV networks
	NDTRP	GPSR [44]	Greedy forwarding: to determine the neighbor closest to the destination and forward the packet accordingly. Perimeter forwarding: to guide the packet around obstacles	When the destination node is located within a “hole”—an area surrounded by nodes that are all farther away from the destination than the source node, greedy forwarding fails
		GLSR [45]	Can handle dynamic UAV networks efficiently, and guarantees loop-free routes by using location information & sequence numbers	Relies on the accuracy of GPS or other positioning systems used by UAVs. Inaccurate location information can lead to suboptimal route calculations and potential routing errors
	DTRP	USMP [46]	Send messages directly to neighboring nodes sharing GPSR location details	Enhance the delivery ratio, reduce the traveled distance
		TENSR [47]	Incorporate “social ties” between UAVs can lead to faster route discovery, improved network efficiency	Scalability challenges, enhance delivery ratio and delay
	HRP	DPTR [48]	Enhance throughput and delay [49]	No consideration of ground mobility [49]

(continued)

**Table 3** (continued)

Category	Sub-category	Example of protocols	Advantages	Limitations
		UVAR [50]	Fully exploits UAVs [49]	Cause high delay, energy overuse [49]
Swarm based		HGA [51]	HGAs can be used to optimize various aspects of UAV networks, such as routing protocols, resource allocation (battery power, bandwidth), and path planning	The effectiveness of HGAs heavily relies on proper parameter tuning, which can be a complex and time-consuming process

## 6 Open Issues and Future Directions

### 6.1 Open Issues

Despite numerous research efforts to enhance UAV networks' efficiency, there remain unresolved issues that need further investigation in future studies. Here, we outline various concerns about how the components of multi-UAV systems are integrated and interact. Our current research primarily addresses open issues related to cooperative target tracking using multiple UAVs, particularly from the perspectives of remote sensing and coordination.

- (a) *UAV Platform*: The flight dynamics platforms of UAVs, such as their vulnerability to weather conditions and wind, as well as the orientation and position of UAVs, have a major impact on the quality of communication links [49].
- (b) *Data Processing*: Data for UAVs is provided through the onboard sensors that are used in the process. Inboard UAVs might not have adequate processing ability to perform the processing of the acquired data. Cloud computing can be mentioned as an alternative that can serve as a solution to this problem.
- (c) *Routing*: In UAV networks, the configuration type is dynamic which means that the structure of the network may change frequently. The management of data exchange between UAVs is considered to be a considerable challenge. Every routing protocol should cause the routing tables to be updated whenever there is a change in the network topology [22].
- (d) *Path Planning*: Proper coordination of the UAVs is paramount when it comes to the use of multiple UAVs within the system especially when it comes to missions. For the coordination of UAVs, the development of new algorithms on dynamic path planning is sufficiently compulsory.

- (e) *QoS Provisioning:* Since UAVs' applications include sending GPS coordinates and video/voice streaming, the UAV network has to provide Quality of Service (QoS) to correspond to a set of high-demand requirements for the services, which consist of low packet loss rate, minimal delay, and adequate bandwidth.
- (f) *Integration of UAVs into Existing Systems:* The common issue for multi-UAV systems is that they can be integrated into existing infrastructure restricted to the scalability problem, the amount of data that is being produced in large-scale systems, as well as the data processing of the data from various devices.

## 6.2 *AI for the Future UAV Communication Systems*

New trends in communication UAV systems will have frequent use of artificial intelligence techniques such as machine learning (ML), deep learning (DL), and federated learning (FL) in the next decade. To improve the UAV communication networks, researchers are continuously trying to execute AI-driven ML and DL models for not only recognizing various types of UAVs but also recognizing and distinguishing their nature of operations. Issues are present in integrating AI in concerns such as position confirmation of UAVs as well as the route and success of a SAR operation. The nature of UAV communication systems is relatively more complex than that of the already established terrestrial networks, and thus researchers have to spend more time identifying which of the explained AI techniques is most suitable for the particular system. Since the ubiquity of AI techniques necessitates more reckoning than conventional methods, computation efficiency has to be optimized for enhancing the communication network of UAVs.

## 6.3 *Identifying UAVs via Radio Signal Analysis*

From the analysis of this survey, there are some following directions that may help in localizing UAVs from the radio frequency (RF) signal analysis [52–54].

- (a) *Frequency Hopping vs. Dedicated Channels:* By analyzing the developing communication protocols of UAVs information about the type of the UAV may be obtained. Whereas, the commercial UAVs fly on pre-designated WiFi or cellular frequencies, and the military, or custom UAVs may use the frequency hopping method to evade identification. Specifying the used communication method allows for the exclusion of a certain type of UAV source.
- (b) *Signal Modulation Techniques:* There are differences in the use of various types of modulation techniques in data transmission of the UAV systems. This is because the modulated signals' type (FSK, QAM) of the intercepted signal can be compared with known UAV communication protocols helping in UAV identification.

- (c) *Packet Analysis:* UAV communication particularly includes packet structures with control information telemetry data as well as very possibly video/audio stream. Analyzing these packets will let you get data on the current mode of the UAV, or sensor data if any, or the manufacturer's unique signatures that are implanted in these packets.
- (d) *Traffic Patterns:* Based on the analysis of the radio traffic characteristics, one can find out more features of flying arouses communication. This might incorporate certain data transmission frequencies, traffic bursts revealed with control commands, or video streaming traffic which is unlikely to be similar to that in the ground-based traffic.
- (e) *Direction Finding:* The use of multiple radio receivers makes it possible for the source of the signal to be triangulated. This technique can locate the approximate position of the UAV; hence, the identification process can be done either visually or using other signals to increase the chances of discovering the UAV.

## 6.4 Recommendations on Future UAV Standards

IEEE is now engaged in the development of many future standards for unmanned aerial vehicles (UAVs).

- (1) IEEE P1920.1 [55] is a set of standards that will establish guidelines for aerial communications and networks [49]. These standards will specifically focus on air-to-air communications inside self-organized ad hoc aerial networks [49].
- (2) Once the IEEE P1920.1 [55] standard is accomplished, the efforts pertaining to the IEEE P1920.2 [56] standards for vehicle-to-vehicle communications for Unmanned Aircraft Systems (UASs) [49] are anticipated to commence.
- (3) The IEEE P1936.1 [57] Standard for Drone Applications Framework (SDAF) attempts to create a framework for supporting drone applications and defining the classes, scenarios, and execution environments necessary for these applications [49].
- (4) The goal of the IEEE P1939.1 [58] initiative is to provide a standardized framework for organizing the airspace at low altitudes for unmanned aerial vehicles (UAVs).
- (5) The IEEE P1937.1 standard outlines the interface requirements and performance characteristics of payload devices in drones, often known as IPDDs [59]. It describes the interfaces, provisioning, performance metrics, operation control, and administration of UAV payload devices [49].

The recommendations described above would necessitate the establishment of a set of novel global standards and widely accepted practices, as well as smooth and easily understandable interfaces and procedures to ensure user-friendliness, extending to a high degree of friendliness and approachability.

## 7 Conclusion

The discussion on the different types of UAV networking models like the multi-UAV or cooperative systems can assist in controlling processes, especially in identification or detection within the environment. On many occasions, multi-UAV systems demand efficient and accurate routing, as well as coordinating procedures in several instances, some of the routing protocols in the classification can work. Perhaps, some other distinctions might include the differences in special operational characteristics such as the operating environment and the coverage, the latency–reliability trade-off. Comparing ZRP to the family of routing protocols, it fits within multi-UAV systems particularly when it facilitates the provision of intervention over large geographical zones. To this end, it uses active routing in the local areas or areas while using passive routing in the areas, which makes it efficient in routing and reducing on expenses. For the systems which can imply a number of cooperative UAVs, HGA can be effective because for instance using evolutionary algorithms the proper paths from the origin point to the goal point can be identified by serving several characteristics such as energy and time requirements. This is especially beneficial in the instances in which there is a requirement to make complex decisions concerning enhancing the motion dynamics of the swarm.

The optimal combination of drones, applications, and communication technology will lead to safer, more dependable, and more efficient drones with longer flight times and minimal communication delays. However, the quick movement and frequent network topology changes of UAVs may create interrupted swarm communication, which has always been a major concern for routing protocols. Therefore, finding a solution to intermittent connections will remain a research priority.

## References

1. Rahman MA (2014) Enabling drone communications with WiMAX technology. In: IISA 2014, the 5th international conference on information, intelligence, systems and applications. IEEE, pp 323–328
2. Yang P, Cao X, Yin C, Xiao Z, Xi X, Wu D (2017) Routing protocol design for drone-cell communication networks. In: 2017 IEEE international conference on communications (ICC). IEEE, pp 1–6
3. Kitagawa T, Ala S, Eum S, Murata M (2018) Mobility-controlled flying routers for information-centric networking. In: 2018 15th IEEE Annual consumer communications & networking conference (CCNC) . IEEE, pp 1–2
4. Yoshikawa K, Yamashita S, Yamamoto K, Nishio T, Morikura M (2017) Resource allocation for 3d drone networks sharing spectrum bands. In: 2017 IEEE 86th vehicular technology conference (VTC-Fall). IEEE, pp. 1–5
5. Sharma A, Vanjani P, Paliwal N, Basnayaka CMW, Jayakody DNK, Wang HC, Muthuchidambaranathan P (2020) Communication and networking technologies for UAVs: a survey. J Netw Comput Appl 168:102739
6. Zhao N, Yang X, Ren A, Zhang Z, Zhao W, Hu F, Ur Rehman M, Abbas H, Abolhasan M (2018) Antenna and propagation considerations for amateur UAV monitoring. IEEE Access 6:28001–28007

7. Multerer T, Ganis A, Prechtel U, Miralles E, Meusling A, Mietzner J, Vossiek M, Loghi M, Ziegler V (2017) Low-cost jamming system against small drones using a 3d MIMO radar-based tracking. In: Radar conference (EURAD), 2017 European. IEEE, pp 299–302
8. Deruyck M, Wyckmans J, Joseph W, Martens L (2018) Designing UAV-aided emergency networks for large-scale disaster scenarios. *EURASIP J Wirel Commun Netw* 2018(1):1–2
9. Zahariadis T, Voulkidis A, Karkazis P, Trakadas P (2017) Preventive maintenance of critical infrastructures using 5G networks & drones. In: 2017 14th IEEE international conference on advanced video and signal based surveillance (AVSS). IEEE, pp 1–4
10. Moon H, Kim C, Lee W (2016) A UAV based 3-d positioning framework for detecting locations of buried persons in collapsed disaster area. *Int Arch Photogramm, Remote Sens Spat Inf Sci* 41:121–4
11. Hayat S, Yanmaz E, Muzaffar R (2016) Survey on unmanned aerial vehicle networks for civil applications: a communications viewpoint. *IEEE Commun Surv Tutor* 18(4):2624–2661
12. He D, Chan S, Guizani M (2016) Communication security of unmanned aerial vehicles. *IEEE Wireless Commun* 24(4):134–139
13. Kanistras K, Martins G, Rutherford MJ, Valavanis KP (2013) A survey of unmanned aerial vehicles (UAVs) for traffic monitoring. In: 2013 International conference on unmanned aircraft systems (ICUAS). IEEE, pp 221–234
14. Sharma V, Kumar R (2017) Cooperative frameworks and network models for flying ad hoc networks: a survey. *Concurr Comput* 29(4):1–36
15. Zeng Y, Zhang R, Lim TJ (2016) Wireless communications with unmanned aerial vehicles: opportunities and challenges. *IEEE Commun Mag* 54(5):36–42
16. Jawhar I, Mohamed N, Al-Jaroodi J, Agrawal DP, Zhang S (2017) Communication and networking of UAV-based systems: classification and associated architectures. *J Netw Comput Appl* 84:93–108
17. Sánchez-García J, García-Campos J, Arzamendia M, Reina DG, Toral S, Gregor D (2018) A survey on unmanned aerial and aquatic vehicle multi-hop networks: wireless communications, evaluation tools and applications. *Comput Commun* 119:43–65
18. Oubbat OS, Atiquzzaman M, Lorenz P, Tareque MH, Hossain MS (2019) Routing in flying ad hoc networks: survey, constraints, and future challenge perspectives. *IEEE Access* 7:81057–81105
19. Chriki A, Touati H, Snoussi H, Kamoun F (2019) FANET: Communication, mobility models and security issues. *Comput Netw* 163:106877
20. Nawaz H, Ali HM, Laghari AA (2020) UAV communication networks issues: a review. *Arch Comput Methods Eng* 1–2
21. Tsao KY, Girdler T, Vassilakis VG (2022) A survey of cyber security threats and solutions for UAV communications and flying ad-hoc networks. *Ad Hoc Netw* 133:102894
22. Chen X, Tang J, Lao S (2020) Review of unmanned aerial vehicle swarm communication architectures and routing protocols. *Appl Sci* 10:3661. <https://doi.org/10.3390/app10103661>
23. Erdelj M, Natalizio E, Chowdhury KR, Akyildiz IF (2017) Help from the sky: leveraging UAVs for disaster management. *IEEE Pervasive Comput* 16(1):24–32
24. Wu T, Yang P, Yan Y, Rao X, Li P, Xu W (2017) ORSCA: optimal route selection and communication association for drones in WSNs. In: 2017 Fifth international conference on advanced cloud and big data (CBD) (pp. 420–424). IEEE
25. Alnoman A, Anpalagan A (2017) On D2D communications for public safety applications. In: Humanitarian technology conference (IHTC), 2017 IEEE Canada international. IEEE, pp 124–127
26. Shi W, Zhou H, Li J, Xu W, Zhang N, Shen X (2018) Drone assisted vehicular networks: Architecture, challenges and opportunities. *IEEE Network*
27. Oubbat OS, Lakas A, Zhou, G'üneş M, Lagraa N, Yagoubi MB (2017) Intelligent UAV-assisted routing protocol for urban VANETs. *Comput Commun* 107:93–111
28. Wang X, Fu L, Zhang Y, Gan X, Wang X (2016) VDNet: an infrastructure-less UAV-assisted sparse VANET system with vehicle location prediction. *Wirel Commun Mob Comput* 16(17):2991–3003

29. Li X, Guo D, Yin H, Wei G (2015) Drone-assisted public safety wireless broadband network. In: Wireless communications and networking conference workshops (WCNCW), 2015. IEEE, pp 323–328
30. Wzorek M, Landén D, Doherty P (2006) GSM technology as a communication media for an autonomous unmanned aerial vehicle. In: Proceedings of the 21st Bristol international conference on UAV systems. Citeseer
31. Muruganathan SD, Lin X, Maattanen H-L, Zou Z, Hapsari WA, Yasukawa S (2018) An overview of 3GPP release-15 study on enhanced lte support for connected drones. arXiv preprint [arXiv:1805.00826](https://arxiv.org/abs/1805.00826)
32. Korhonen J (2018) Enhanced lte support for aerial vehicles. 3rd Generation partnership project (3GPP), Tech. Rep. [Online]. Available: <https://www.3gpp.org/ftp/Specs/archive/36series/36.777/>
33. Chandhar P, Danev D, Larsson EG (2017) Massive MIMO for communications with drone swarms. *IEEE Trans Wireless Commun* 17(3):1604–1629
34. Pandey GK, Gurjar DS, Nguyen HH, Yadav S (2022) Security threats and mitigation techniques in UAV communications: a comprehensive survey. *IEEE Access* 10:112858–112897
35. Bekmezci I, Sahingoz OK, Temel S (2013) Flying ad-hoc networks (FANETs): a survey. *Ad Hoc Netw* 11(3):1254–1270
36. Sahingoz OK (2014) Networking models in flying ad-hoc networks (FANETs): concepts and challenges. *J Intell Rob Syst* 74(1–2):513–527
37. Cheng C-M, Hsiao P-H, Kung H, Vlah D (2007) Maximizing throughput of UAV-relaying networks with the load-carry-and-deliver paradigm. In: 2007 IEEE wireless communications and networking conference, IEEE, pp 4417–4424
38. Jacquet P, Muhlethaler P, Clausen T, Laouiti A, Qayyum A, Viennot L (2001) Optimized link state routing protocol for ad hoc networks, proceedings. In: IEEE International multi topic conference (INMIC), 2001. Technology for the 21st century. IEEE, pp 62–68
39. He G (2002) Destination-sequenced distance vector (DSDV) protocol. Helsinki University of Technology, Networking Laboratory, pp 1–9
40. Johnson DB, Maltz DA (1996) Dynamic source routing in ad hoc wireless networks. Springer, Mobile Computing, pp 153–181
41. Chakeres ID, Belding-Royer EM (2004) AODV routing protocol implementation design. In: 24th international conference on distributed computing systems workshops. Proceedings. IEEE, pp 698–703
42. Beijar N (2002) Zone routing protocol (ZRP), 9 networking laboratory. Helsinki University of Technology, Finland, pp 1–12
43. Park V, Corson S (2001) Temporally-ordered routing algorithm (TORA) version 1, Functional specific. IETF Internet Draft
44. Karp B, Kung H-T (2000) GPSR: greedy perimeter stateless routing for wireless networks. In: Proceedings of the 6th annual international conference on mobile computing and networking, pp 243–254
45. Medina D, Hoffmann F, Rossetto F, Rokitansky C-H (2010) Routing in the airborne internet. In: 2010 Integrated communications, navigation, and surveillance conference proceedings. IEEE, pp A7–1–A7–10
46. Lidowski RL, Mullins BE, Baldwin RO (2009) A novel communications protocol using geographic routing for swarming UAVS performing a search mission. In: 2009 IEEE international conference on pervasive computing and communications. IEEE, pp 7–10
47. Gupta R, Krishnamurthi N, Wang U-T, Tammineni T, Gerla M (2017) Routing in mobile ad-hoc networks using social tie strengths and mobility plans. In: 2017 IEEE wireless communications and networking conference (WCNC). IEEE, pp 1–6
48. Sharma V, Kumar R, Kumar N (2018) DPTR: distributed priority tree-based routing protocol for FANETs. *Comput Commun* 122:129–151
49. Hentati AI, Fourati LC (2020) Comprehensive survey of UAVs communication networks. *Comput Stand Interfaces* 72:103451

50. Oubbat OS, Lakas A, Lagraa N, Yagoubi MB (2016) UVAR: an intersection UAV-assisted VANET routing protocol. In: 2016 IEEE wireless communications and networking conference. IEEE, pp 1–6
51. Peng K, Du J, Lu F, Sun Q, Dong Y, Zhou P, Hu M (2019) A hybrid genetic algorithm on routing and scheduling for vehicle-assisted multi-drone parcel delivery. *IEEE Access* 7:49191–49200
52. Podder P, Zawodniok M, Madria S (2024) Deep learning for UAV detection and classification via Radio frequency signal analysis. In: 2024 25th IEEE international conference on mobile data management (MDM), Brussels, Belgium, pp 165–174, <https://doi.org/10.1109/MDM61037.2024.00040>
53. Liu Y, Qin Z, Cai Y, Gao Y, Li GY, Nallanathan A (2019) UAV communications based on nonorthogonal multiple access. *IEEE Wirel Commun* 26(1):52–57
54. Yan L (2024) UAV detection with radio frequency data and deep learning techniques. In: Proceedings of 2024 IEEE 6th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), pp 234–239
55. IEEE, P1920.1: standards for aerial communications and networks. [https://standards.ieee.org/project/1920\\_1.html](https://standards.ieee.org/project/1920_1.html)
56. IEEE, P1920.2: standards for vehicle to vehicle communications for UAS (unmanned aircraft systems). [https://standards.ieee.org/project/1920\\_2.html](https://standards.ieee.org/project/1920_2.html)
57. IEEE, P1936.1 standard for drone applications framework (SDAF). [https://standards.ieee.org/project/1936\\_1.htmlc](https://standards.ieee.org/project/1936_1.htmlc)
58. IEEE, P1939.1 standard for a framework for structuring low altitude airspace for unmanned aerial vehicle (UAV) operations. [https://standards.ieee.org/project/1939\\_1.htmld](https://standards.ieee.org/project/1939_1.htmld)
59. IEEE P1937.1 standard interface requirements and performance characteristics of payload devices in drones (IPDD). [https://standards.ieee.org/project/1937\\_1.html](https://standards.ieee.org/project/1937_1.html)

# Securing Patient Personal Information Using Multi-Dimensional Anonymization-Based Intelligent Technology Using Edge Nodes



Abhinav Yadav, Marushika Shukla, Tiansheng Yang,  
Rajkumar Singh Rathore, and Hrudaya Kumar Tripathy

**Abstract** Ensuring patient privacy is a primary concern when considering the integration of artificial intelligence in healthcare. Advanced models have the capability to utilize and safeguard diverse patient datasets, ensuring secure data exchange and concealing personal health information. They can adapt to enhance blended learning, blockchain, natural language processing, cybersecurity, biometric authentication, and other techniques. However, ethical considerations, such as defining limits and eliminating biases, pose significant challenges. To address these concerns, increasing transparency and minimizing prejudice are crucial steps for the ethical integration of AI. In summary, the adoption of artificial intelligence in healthcare presents a significant opportunity to enhance patient privacy by implementing safeguard measures against unauthorized access to private information.

**Keywords** Anonymization · Edge computing · Patient security · Healthcare · Blockchain

---

A. Yadav · M. Shukla · H. K. Tripathy

Kalinga Institute of Industrial Technology, Deemed to Be University, Bhubaneswar, India  
e-mail: [hktripathyfcs@kiit.ac.in](mailto:hktripathyfcs@kiit.ac.in)

T. Yang (✉)

University of South Wales, Pontypridd, UK  
e-mail: [tiansheng.yang1@southwales.ac.uk](mailto:tiansheng.yang1@southwales.ac.uk)

R. S. Rathore

Cardiff School of Technologies, Cardiff Metropolitan University, Llandaff Campus, Cardiff, UK  
e-mail: [rsrathore@cardiffmet.ac.uk](mailto:rsrathore@cardiffmet.ac.uk)

## 1 Introduction

The World Health Organization (WHO) recognizes that AI has an immense capability to enhance healthcare and medicine delivery [1]. AI Technology is currently being used in medicine for a variety of purposes, including drug discovery, medical imaging, personalized disease therapy, and disease detection and diagnosis. AI can help with the implementation of interventions like disease surveillance, outbreak response, and healthcare system management in the domains of epidemiology and public health. Human rights and medical ethics are simultaneously jeopardized by the applications of AI. AI has the potential to compromise human autonomy and dignity in decision-making, compromise personal privacy, and lead to algorithmic discrimination, among other negative effects. The issues of protecting patient privacy rights have been a lingering concern in recent years with the use of AI in healthcare. When the patient privacy is violated, negative consequences include employment discrimination and increased long-term healthcare costs [2]. Healthcare data was recorded and stored on physical documents traditionally. Medical institutions ensured the protection of patient confidentiality among their staff. In today's healthcare landscape, patient data is frequently collected and sent digitally to a larger, more fluid healthcare data pool for various uses. Privacy issues are becoming increasingly complex in the era of convenient information sharing due to the collection and use of patient data in AI-related medical technology. The research was conducted extensively to gain a better understanding of the expectations and behaviors of a diverse range of users, with the primary goal being that the application be simple to use. Maintaining an engaging interface, the interaction remains lively. The application's operation will appear simple and straightforward to the end user due to its design. Users can be categorized into two groups based on their comprehension of the goods that best suit their needs. The two groups are formed by the people: one consists of those who are already acquainted with the product that caters to their requirements, while the other is still in the process of selecting a product that will suit their needs. A single click of a button is all it takes for users who are already familiar with the product to find it quickly and easily. The product can be found by users in this category through a browser search, using the product name as the query[3]. Various criteria are used to filter the results of a keyword search for products. The product specification and multiple images taken at various zoom levels should be visible to buyers. The following information should be provided to enable the user to see the product, in addition to customer reviews and ratings. They should be able to write their own reviews if they choose to[4]. A product's specifications, for instance, should be provided in print, and the product page should be sharable with friends. The shopping cart should be made to accept products when they are dragged and dropped into it for an enhanced user experience. The shopping cart should allow users to make adjustments to their orders. They can modify the quantities of added products to remove items as needed. The users can remove a product from their shopping cart by pulling it out and dropping it outside the cart's boundaries. The software could be improved by using pop-up messages. They could make the experience more interactive by displaying when a product is

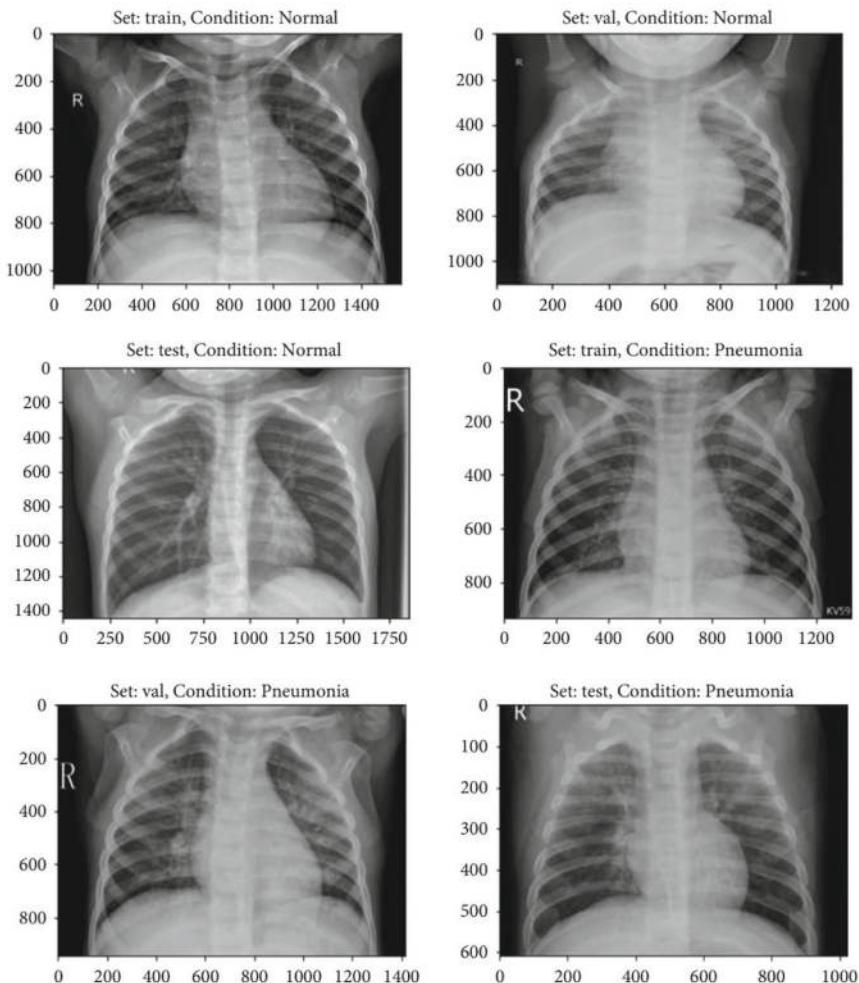
added to or removed from the shopping cart. When a user arrives at a drop location, they will be informed and can detect the likely item to be dropped. People's patience is limited, so websites must load quickly to be successful.

## 2 Literature Review

A variety of methodologies have been used by the authors in their current approaches. The shading highlights extracted from the spatially organized L2 standardized coefficients have been coupled with the processed form and item data to create a more realistic appearance. In portion of the current techniques, the SIFT (scale-invariant feature transform) process is used to locate and depict highlights in computer-generated images. The application can identify the specific core concerns and provide them with quantitative data, referred to as descriptors, for use in object recognition [5]. More effective search strategies must be used when dealing with massive amounts of data instead of a brute force search. A small number of visual word histograms represent a large number of pictures in a simple encoding system called the bag of features, using a limited amount of space [6]. This inverted index data structure allows for small storage and excellent search performance. The type of items in the database influences what is used in retrieval functions. The object's characteristics are extracted from its surroundings using a function bag technique [7]. An object is calculated, generating a dictionary after numerous representative keywords have been excluded, based. A large amount of data is required to develop a robust vocabulary, (which means a large dataset)[8]. The bag of an item can be found by searching for the most linked cluster centers in the dictionary for each of its features. The descriptor space changes and maps each item's features and descriptors, frequently using SIFT. The SIFT functions in the image can be used to search the image database for the most comparable visual word in the lexicon for each one. This can be expressed in the passive voice as: The most comparable visual word in the lexicon for each SIFT function in the image can be discovered by searching the image database. A k-dimensional histogram of the SIFT functions for each item in the dictionary can be created to count the items [9]. Online purchasing activity has surged hugely in recent months due to the COVID-19 epidemic suggesting the technique. Consumer service awareness is moderated by COVID-19, enabling shoppers to acquire products and services online [10]. Use the Bag of Features technique to extract identical images in the Healthcare Monitoring service, as proposed. The query object, as shown in Fig. 1, is retrieved by indexing an object that has been previously trained as part of a Bag of Features, which is generated by training a succession of random things from the dataset. The limitations in terms of accuracy and issues with the datasets used in the development of preceding image retrieval systems have been previously noted. Can be converted to: Image retrieval systems based on preceding technologies have limitations in accuracy and face challenges with the datasets used in their development [11]. The datasets are increased in quality and modified to surpass the limitations of previous articles. The processes for obtaining comparable things for purchasing are

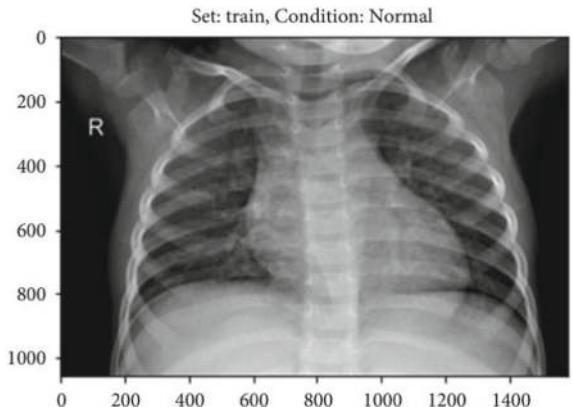
as follows: (A) choose the item to be retrieved; (B) generate a bag of characteristics for the object to be recovered; (C) categorize the item; and (D) search the index for comparable objects. Figure 2 shows the output scan of the sample query object.

- (A) Choose the Item to be retrieved. An image collection that includes scenery (beaches, towns, roads, etc.) is what one looks for, the color contents of the entire scene, as captured in a global object characteristic such as a color histogram, are preferable over more particular object characteristics. The essentials objects in a sequence of photographs are more effectively located by utilizing local image characteristics around them [12].



**Fig. 1** Query object

**Fig. 2** Output for the given query object



- (B) A bag of feature is formed by selecting a random sample of objects from the dataset and training them, after which the features are put into the bag. In a Bag of Features approach, labels are not necessary for training the feature extraction process. This learning technique is classified as weakly supervised since it does not rely on a substantial number of labels. This approach, on the other hand, does not take into account the spatial link between features in any way. A little higher score causes discomfort when there are multiple good potential partners to choose from. In this situation, a mechanism needs to be provided that accurately distinguishes among several options to identify the best one. Including geometric information alongside the text can help prevent this issue.
- (C) An adequate dataset and a substantial quantity of data are required to build a suitable vocabulary for creating an image index of all the images [13]. In this process, the descriptor space is mapped, and the attributes and descriptors for each item are displayed. This is commonly how descriptors are represented in the descriptor space. Visual words within a query object must be determined before searching for them. The phrase being looked for in an item is checked to see if it appears in any other items [14]. One vote is added to the number of votes for each item in the voting array. Each item in the array is a list with a counter variable for the object it represents. Example: [3, 5, 7] - Five votes for the first object, three votes for the second object, and seven votes for the third object. The query object's match is selected from the voting array with the highest counter value at the operation's end. Every visual word in the vocabulary must be compared must be compared to each property in the query object, even if the query object contains no visual words.
- (D) You should consider items that are similar to yours. The next step is to use the get objects method to identify comparable items. In passive voice: Items that are similar to yours should be considered. The next step is to have the get objects method identify comparable items. The user must choose a photo query (input object) for the result. To obtain photographs from the dataset, he uses the query picture as his user image [15]. The image is not required to be from

our dataset for the picture query to be valid, and there are no source limitations. The computer takes a snapshot of the query image displayed in Fig. 1 as soon as it is submitted into it.

Step 1: The set of training products is used to learn the visual vocabulary included within the bag of features after defining the feature type. The “Custom Extractor” option extracts features from each product and generates a bag of them by randomly selecting a collection of items from the dataset. Now that the feature bag has been produced, the whole X-ray picture set may be indexed for searching [16].

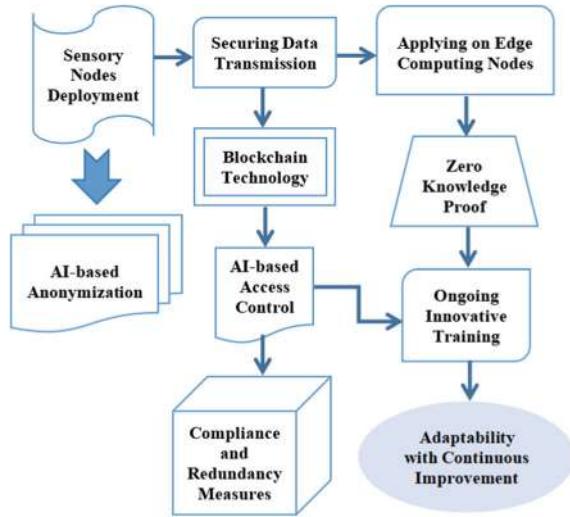
Step 2: The X-ray image collection may be indexed for searches before the formation of feature bag. Using the step 1 tailored extractor function, the indexing process extracts characteristics from each product.

Step 3: The fetch photos tool is used to look for comparable photographs. In Fig. 2, output 1 met the accuracy of objects.

### 3 Proposed Model and Architecture

The following documentation explains a multi-dimensional approach that was implemented to create security of patient data in healthcare domain using Internet of Things [17–19]. The model includes strategic aspects, starting from IoT devices deployment comprised of such elements as sensors and medical equipment bearing in mind security standards and data and communication protocols security, which guarantees integrity and privacy of the data. Edge nodes are provided in healthcare centers for the purpose of local processing of data thus reducing latency and the need to send sensitive data over networks. Major attention is on employing relied encryption technologies including MQTT, TLS, and HTTPS to encrypt both data transit as well as data at rest. Besides that, the adoption of AI-enabled anonymization modules achieves identification of patient data on the edge, meanwhile the adoption of various privacy mechanisms also contributes to the thorough protection from possible re-identification attacks. Blockchain technology in medical field is deployed to check the accuracy and immutability of patient data and also provide a secure and tamper-proof record. Furthermore, zero-knowledge proofs do not allow the arithmetic computation of encrypted data that may disclose the sensitive data [1, 20–22]. The AI-powered access control system dynamic adapts the access rights based on the user roles and behavior and reinforces with behavioral analytics which detects the anomalies and enabling the prevention to security threats. Furthermore, the continuous improvement of AI models and security provisions, alongside extensive user training, creates a flexibly-adaptive environment to emerging privacy problems and enables the medical community to be aware of privacy regulations [23]. Moreover, compliance mechanisms and disaster recovery plans that are difficult to manipulate and well-structured information security architecture maintain an unwavering access to the crucial patient information in healthcare facilities.

**Fig. 3** Proposed multi-dimensional edge computing model



In the given Fig. 3, we can see the work flow of the proposed model that how the data is traveling through each node and how it is getting secured.

## 4 Result and Analysis

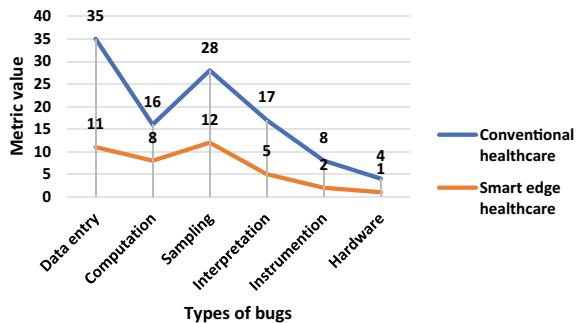
The presented observation in Table 1 highlights the efficacy of employing a smart method, which has significantly reduced the time required for obtaining results compared to conventional approaches. This suggests a clear advantage in the effectiveness of the smart method, thereby advocating for the integration of new technologies into the treatment process.

The reduction in the number of days through the smart method offers several noteworthy benefits:

**Table 1** Computational delay analysis using conventional healthcare and smart healthcare

Health risks	Traditional approach	Smart healthcare
Hepatitis	30 Days	20 Days
Malaria	7 Days	4 Days
Breast cancer	40 Days	33 Days
Dengue	14 Days	10 Days
Cholera	5 Days	2 Days
Typhoid	14 Days	8 Days
Mean computation delay	18.33 Days	12.83 Days

**Fig. 4** Bugs repeat value analysis for both conventional and smart edge healthcare



- (a) Accelerated Recovery: The smart method expedites patients' healing processes, leading to improved health outcomes and enhanced quality of life [24].
- (b) Cost-Efficiency: A shortened treatment duration translates to reduced healthcare costs, encompassing expenses related to hospital stays, medications, and other healthcare services.
- (c) Enhanced Patient Experience: A concise therapy course contributes to heightened patient satisfaction and minimized side effects [25, 26].
- (d) Optimal Resource Utilization: The adoption of the smart method proves advantageous for healthcare facilities, promoting resource optimization and increased productivity.

The experimentally proven smart edge approach was with a step-by-step progress in the performance, that is, a lower error rate as compared to the traditional approach, which highlights a significant advantage in terms of accuracy and also a positive impact on the reliability of smart method. Figure 4 highlights the bugs repeat metric analysis for both traditional as well as smart edge healthcare. Here are the implications of this observation: It follows this evidence, various implications are discussed below:

- (a) Data Entry Errors: It gives the guarantee of the reliability of the data used in the pursuing the treatment steps, hence minimizing the risk of patients' data errors.
- (b) Calculation Errors: One cannot ignore the role of precision in computations needed to rise the significance of the therapeutic plans.
- (c) Sampling Errors: In the conclusion, the writing is also better which makes the trustworthiness of the results on the analysis reinforced.
- (d) Interpretation Errors: There will be less misunderstandings, thus more accurate therapy assessment and recommendations should be made, by which more personalized treatment approaches can be applied for better patient's outcomes.
- (e) Instrumentation Errors: Development of humanized diagnostic test accuracy will eliminate danger from misdiagnosis or error in the delivery of treatment.
- (f) Reporting Errors: The machine-learning tools employed into the system enable the system to eliminate the errors that surface in the reporting of the treatment's outcome.

## 5 Conclusion

After the AI models are applied in healthcare, the privacy of the patients may be a little bit modified. As the healthcare providers can always secure the sensitive patient data from any illegal access while yet allowing authorized internal access by the health care professionals for research and analysis by adding AI models into their operations. AI models can also be used to design consented data-sharing platforms, detect potential breaches, catch patient privacy violations in real time and to build consent management systems. Nevertheless, the AI models can act as an immensely helpful tool to provide patients' privilege, but they are not a total solution. Besides, the privacy and security aspects that healthcare providers should bring up as critical include staff training and education, risk assessment and management, and employment of sophisticated security methods and technology. Therefore, an all-encompassing strategy which includes AI models while taking into account the extra precautions for privacy and security, is needed to preserve patient privacy in healthcare settings. While AI modeling is being applied in healthcare decision-making, it should be done through prudence. Healthcare AI models may raise moral concerns due to the utilization due to various reasons like biases in data and decision-making processes. Hence, it is important to make sure that the AI models are straightforward, understandable, and free from any bias.

## References

1. Mishra S, Chaudhury P, Tripathy HK, Sahoo KS, Jhanji NZ, Hassan Elnour AA, Abdelmaboud A (2024) Enhancing health care through medical cognitive virtual agents. *Digital Health* 10:20552076241256732
2. Jain S, Tripathy HK (2024) Machine learning methods for the timely identification of autism spectrum disorder in toddlers. In: 2024 International conference on emerging systems and intelligent computing (ESIC). IEEE, pp 515–520
3. Kashyap P, Pareek A, Mishra S, Khan Z, Garg R, Tripathy HK (2024) Sentiment polarity analysis of twitter data using machine learning models. In: International conference on innovative computing and communication. Springer Nature Singapore, Singapore, pp 623–635
4. Li Z, Wei CH, Li Y, Sun T (2011) Research of shoeprint image stream retrieval algorithm with scale-invariance feature transform. In: 2011 International conference on multimedia technology, pp 5488–5491
5. Parizi SN, Laptev I, Targhi AT (2009) Modeling image context using object centered grid. In: 2009 Digital image computing: techniques and applications, pp 476–483
6. Acton, ST, Rossi A (2008) Matching and retrieval of tattoo images: active contour CBIR and glocal image features. In: 2008 IEEE southwest symposium on image analysis and interpretation, pp 21–24
7. Zhang R, Zhang Z (2007) Effective image retrieval based on hidden concept discovery in image database. *IEEE Trans Image Process* 16:562–572
8. Kumar T, Sreenivasa Murthy A, Rajani N (2016) HPCIR: histogram positional centroid for image retrieval. In: 2016 IEEE industrial electronics and applications conference (IEACon), pp 256–260
9. Li D, Luo Z, Cao B (2021) Blockchain-based federated learning methodologies in smart environments. *Clust Comput* 25:2585–2599

10. Kumar S, Singh A, Benslimane A, Chithaluru P, Albahar MA, Rathore RS, Álvarez RM (2023) An optimized intelligent computational security model for interconnected blockchain-IoT system & cities. *Ad Hoc Netw* 151:103299
11. Hassan MM, Zaman S, Rahman MM, Bairagi AK, El-Shafai W, Rathore RS, Gupta D (2024) Efficient prediction of coronary artery disease using machine learning algorithms with feature selection techniques. *Comput Electr Eng* 115:109130
12. Jaime FJ, Muñoz A, Rodríguez-Gómez F, Jerez-Calero A (2023) Strengthening privacy and data security in biomedical microelectromechanical systems by IoT communication security and protection in smart healthcare. *SensS (Basel Switz)* 23(21):8944
13. Sinha P, Sahu D, Prakash S, Yang T, Rathore RS, Pandey VK (2025) A high performance hybrid LSTM CNN secure architecture for IoT environments using deep learning. *Sci Rep* 15(1):9684
14. Reddy S, Allan S, Coghlan S, Cooper P (2020) A governance model for the application of AI in health care. *J Am Med Inform Assoc* 27(3):491–7
15. Stewart KA, Segars AH (2002) An empirical examination of the concern for information privacy instrument. *Inf Syst Res* 13(1):36–49
16. Luxton DD (2019) Should Watson be consulted for a second opinion? *AMA J Ethics* 21(2):131–7
17. Bhawana, Kumar S, Rathore RS, Mahmud M, Kaiwartya O, Lloret J (2022) BEST—blockchain-enabled secure and trusted public emergency services for smart cities environment. *Sensors* 22(15):5733
18. Kumar S, Rathore RS, Dohare U, Kaiwartya O, Lloret J, Kumar N (2023) BEET: blockchain enabled energy trading for e-mobility oriented electric vehicles. *IEEE Trans Mob Comput* 23(4):3018–34
19. Kumar G, Rathore RS, Thakur K, Almadhor A, Biabani SAA, Chander S (2023) Dynamic routing approach for enhancing source location privacy in wireless sensor networks. *Wirel Netw* 29(6):2591–2607
20. Rathore RS, Sangwan S, Kaiwartya O (2021) Towards trusted green computing for wireless sensor networks: multi metric optimization approach. *Adhoc Sens Wirel Netw* 49
21. Saleh A, Joshi P, Rathore RS, Sengar SS (2022) Trust-aware routing mechanism through an edge node for IoT-enabled sensor networks. *Sensors* 22(20):7820
22. Sahoo S, Mishra S, Brahma B, Barsocchi P, Bhoi AK (2024) SSO-CCNN: a correlation-based optimized deep CNN for brain tumor classification using sampled PGGAN. *Int J Comput Intell Syst* 17(1):1–18
23. Mishra S, Jena L, Mishra N, Chang HT (2024) PD-DETECTOR: a sustainable and computationally intelligent mobile application model for Parkinson's disease severity assessment. *Heliyon* 10(14)
24. Pranjal P, Mallick S, Paul A, Mishra S, Bhardwaj I, Albuquerque VHCD (2024) Soil crops and nutrients forecasting using random forest model. In: AIP conference proceedings, vol 2919(1). AIP Publishing
25. Mishra S, Chakraborty S, Sahoo KS, Bilal M (2023) Cogni-Sec: a secure cognitive enabled distributed reinforcement learning model for medical cyber-physical system. *Internet of Things* 24:100978
26. Mishra S, Volety DR, Bohra N, Alfarhood S, Safran M (2023) A smart and sustainable framework for millet crop monitoring equipped with disease detection using enhanced predictive intelligence. *Alex Eng J* 83:298–306

# Tuberculosis Disease Detection: Comparative Analysis of Logistic Regression and Decision Tree Models for Predicting TB Positivity Using Demographic and Symptom Data



Ajay Kumar Tiwari and Alok Katiyar

**Abstract** The present work aims at comparing the role and efficiency of two artificial neural networks, which are logistic regression and decision tree models for the detection of TB positivity depending on demographic factors and symptoms. It has also provided the patient details such as age, gender, and different symptoms of TB. Both were split into training and testing sets with the assessment of the results' accuracy based on the comparison of the results to the data, confusion matrices and ROC curves. Logistic regression model yielded an accuracy of 90% while decision tree yielded an accuracy of 85% as shown by the results above. The ROC curves presented herein showed a good predictive ability of both models where AUC values of 0.85 for the Naive Bayes and 0. Based on the results obtained from the experiments, we have identified that the accuracy achieved for this model is 92 for the decision tree. Even though, in this case the logistic regression's accuracy is slightly higher than the decision tree, both types of models offer a good prognosis for TB diagnostics. The findings of this comparative analysis support the consideration of applying machine learning for the improvement of the accuracy in diagnosing TB and presents future directions for this line of research.

**Keywords** Tuberculosis · Decision trees · Machine learning · Logistic regression · Predictive modeling · Symptom data · Demographic data · Accuracy assessment · Disease detection · ROC analysis

---

A. K. Tiwari (✉) · A. Katiyar

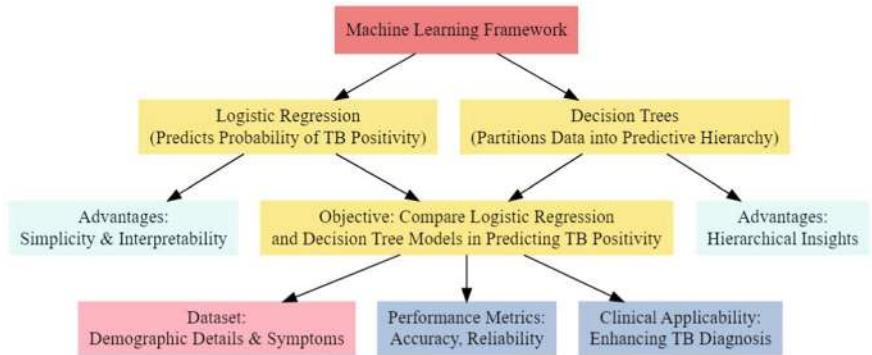
School of Computer Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India

e-mail: [ajaytiwari04@gmail.com](mailto:ajaytiwari04@gmail.com)

## 1 Introduction

Tuberculosis (TB) is a contagious disease that continues to be one of the world's most lethal infectious diseases; new cases are estimated at 10 million and deaths at 1. According to the data available 4 million deaths were recorded in the year 2019 [1]. TB remains a major source of public health concerns and it is still spreading worldwide despite efforts to try and contain it; especially in areas with few health care facilities and high incidence. Fact and accurate diagnosis is essential to not only start an appropriate treatment ASAP but also for preventing the spread of the virus in communities. The conventional microbiological methods of diagnosing TB include sputum smear microscopy and culture which have several limitations with regard to sensitivity, specificity and time taken [2]. These challenges are compounded especially when there is poor health facility endowed and or inaccessible. Recently there has been an increasing focus on applying the ML approaches for improving the diagnosis of TB with the help of the analysis of the patients' data such as the data on the demographic characteristics and the signs of the disease. Machine learning is an approach of great importance in today's world, which enables processing large and comprehensive data, identifying patterns in them and making predictions. From the family of ML algorithms logistic regression and decision trees have come out as very useful in medical diagnosis because of their interpretative capacities and their versatility in handling either categorical or continuous data. Another reason for using logistic regression is because it deals with the probability of getting a binary outcome out of the input variables, thus useful when predicting the probability of getting a TB-positive patient out of the characteristics of the patient [2]. While decision trees split data into subsets based on a hierarchy of values of attributes, this paper applies decision trees to unveil hierarchy of the factors that contribute to the diagnosis of TB [3]. The focus of this research is to analyze the feasibility of utilizing logistic regression and decision tree algorithms in the assessment of TB positivity using a more elaborate data set containing binary information on the patients' demographic characteristics and their symptoms.

The aim is, therefore, to evaluate the validity and reliability of these models of the improvement of TB diagnosis and their usability in the field. The reason for such a comparison is in the possibility of using ML models as supplements or even substitutes for standard approaches to diagnostics. Logistic regression is easy to use and results can be easily interpreted directly which will benefit clinical administrative decision-making that may not be complicated [4, 5]. TB accurately predicts the likelihood of TB positivity depending on certain attributes such as age, gender as well as symptoms hence being able to assist clinicians in arriving at the proper diagnostic decision. Figure 1 exhibits the machine learning framework for processing big data and pays special attention to logistic regression and decision tree model together with their strengths. The aim here is to find out how these models perform in terms of their evaluation metrics and suitability in improving the diagnosis of TB by comparing them with a large data set. The first model of logistic regression aims at analyzing the adjustment probabilities for TB positivity by targeting the characteristics of the



**Fig. 1** Comparative analysis of logistic regression and decision tree models in predicting TB positivity

patients, while the decision tree can give hierarchical identification based on the splitting of data.

However, decision trees seem to be effective in capturing the interactions between predictors making it possibly capable of identifying non-linear relations that would be important in the diagnosis of TB [6]. Decision trees show a clear structure of the hierarchy of rules that allow to understand, which of the demographical and symptom features affect the TB outcomes most. From this study, it is possible to furnish additional real-life information on the applicability of ML models in diagnosis of TB focusing on their advantages and drawbacks [7]. This study aims to compare a logistic regression model to a decision tree model striving to draw conclusions that will help the healthcare practitioners and policymakers to use the ML methods in improving the TB control. In the long run, our target is to prevent and to manage the cases of TB and this can only be achieved if cases are detected early and intervention done appropriately. In conclusion, this paper has given a comparison of logistic regression and decision tree in the classification of TB patients from non-TB patients and illustrated their enormous possibility in shifting the combating systems of tuberculosis in the world.

## 2 Background and Related Work

Tuberculosis continues to be a major threat in the global community especially in the developing nations where there is scarcity of health facilities and diagnostic centers [8]. Tuberculosis occurs due to *Mycobacterium tuberculosis* and mainly affects the pulmonary system. The transmission of TB is through droplets in the air containing the bacteria, and this puts a patient at risk of contracting TB if he or she has close contact with an infectious person for long and uninterrupted periods of time [9].

## **2.1 *Tuberculosis (TB) Epidemiology and Diagnosis Challenges***

In general, identification of TB has been carried out Sameer micro-copy and culture which are microbiological methods. Although these techniques are cheap and readily available, they are less sensitive especially in paucibacillary disease or extrapulmonary tuberculosis [10]. The causes of delay in diagnosis are that the chain of transmission continues and, the disease becomes worse, which demonstrates that there is a need for enhanced diagnostic tests that should be highly sensitive and rapid.

## **2.2 *Role of Machine Learning in TB Diagnosis***

Over the past few years, the application of ML approaches has emerged as a potential means of improving the efficiency of TB diagnosis with the help of the relevant patient data to increase the predictive power of the corresponding algorithms. As it was presented, ML algorithms can solve numerous sophisticated cases that are connected to demographic data, symptoms of the diseases, and laboratory results aiming at the detection of typical indicators for TB infection or disease progression [9, 10]. It can be used as an additional method that is supplementary to the conventional microbiological tests and it also provides possible solutions to the challenges that come with the tests.

## **2.3 *Existing Research on ML Applications in TB Diagnosis***

Some of the researches that have been done concerning the use of ML in diagnosing TB include; For example, one study by Zheng et al., they compared SVM as a reliable model for the classification of TB based on clinical and radiological factors giving high accuracy and sensitivity result [11]. In the same way, using patients' data, AI-enabled study by Khan et al. applied artificial neural networks (ANNs) for determining TB treatment outcomes, ascribing to intricate frameworks of ML great potential in comprehending disease progression [11].

It is noteworthy that logistic regression, and decision trees, as well as SVMs and ANNs, are among the most investigational methods for diagnosing TB. In another study, Chandrasekaran et al., adapted the use of logistic regression and decision tree models wherein the subject of interest was treatment outcome of TB patients based on demographic attributes and co-morbidities [8]. In addition to patient risk evaluation, mentioned models enabled the planning of individual treatment as well.

## 2.4 *Objective of This Study*

Building upon previous research, this study aims to evaluate the performance of logistic regression and decision tree models specifically for predicting TB positivity using demographic and symptom data. By systematically comparing these models, we seek to identify their strengths and limitations in the context of TB diagnosis. This comparative analysis is crucial for informing healthcare practitioners and policymakers about the potential integration of ML-driven approaches into routine TB screening and diagnostic protocols.

## 2.5 *Challenges and Opportunities*

While ML holds promise in improving TB diagnosis, several challenges must be addressed. These include the need for high-quality, standardized datasets, particularly in resource-constrained settings where data collection may be sparse or incomplete. Furthermore, the interpretability of ML models remains a concern, as clinical decisions must be based on transparent and clinically meaningful criteria [9]. Overcoming these challenges presents opportunities to harness the full potential of ML in advancing TB diagnostics and ultimately reducing the global burden of this disease. In summary, this section has provided an overview of TB epidemiology, highlighted the challenges in current diagnostic methods, reviewed existing literature on ML applications in TB diagnosis, and outlined the objectives and rationale for this study. The subsequent sections will delve into the methodology, results, and discussion of the comparative analysis between logistic regression and decision tree models for TB detection.

## 3 Methodology

### 3.1 *Data Collection and Preprocessing*

The dataset used in this study comprises anonymized patient records from [mention source or organization], encompassing demographic information and clinical symptom profiles relevant to TB diagnosis. Variables included age, gender, geographical location, and a comprehensive list of symptoms commonly associated with TB, such as cough, fever, night sweats, weight loss, and hemoptysis. Each patient record was labeled based on microbiological confirmation of TB positivity or negativity using established diagnostic criteria [8, 9].

Prior to model development, data preprocessing was conducted to ensure quality and consistency. This involved handling missing values through imputation techniques suitable for categorical and numerical variables. The categorical variables

were also transformed into a numerical form as required by most of the ML algorithms by operating a one-hot encoding on them. Hierarchical Covariance estimator was applied to the continuous variables and then the continuous variables were transformed to have a zero mean and unit variance in order to reduce the influence of measurement scales of the model performance [10, 12].

### ***3.2 Model Development and Evaluation***

It was decided to select two fundamental ML algorithms for comparison: logistic regression and decision trees [13]. These two algorithms were selected because they are interpretable and have proven to be effective in the field of medical diagnosis, particularly in the area where clear and unambiguous decisions are important.

### ***3.3 Logistic Regression***

Logistic regression is a statistical procedure that predicts the likelihood of a dichotomous outcome such as a TB-positive and negative based on certain independent or input variables. In the present investigation, the logistic regression was applied to the dataset and was used to assess predictive power for each of the three variables under investigation (age, gender, and symptoms) to be able to derive the best model for the probability of being positive to TB. Several metrics such as accuracy, precision, recall, F1-score, AUC-ROC and others were used to evaluate the model performance [2].

### ***3.4 Decision Trees***

For each split decision tree breaks the data into subsets—recursively repeating that process to build branches of leaf nodes until information gain in some criterion is satisfied. The most important demographic and symptom predictors for TB positivity were identified in a decision tree. Some rules were selected and implemented based on the depth, complexity to reach a compromise between interpretability of models and accuracy. Evaluation metrics such as accuracy, precision, recall, F1-score, AUC-ROC were also calculated with respect to decision tree performance [7].

### ***3.5 Model Training and Validation***

To produce the logistic regression and decision tree models, we trained (70%) and validated a data set that was randomly divided into training vs. testing subsets (30%). During the training phase, cross-validation techniques (e.g., K-fold) were used to avoid overfitting and help in obtaining a model that is more robust for predictions on unseen patient data. We used hyperparameter tuning of grid search with cross-validation to optimize model performance, this includes the regularization parameter in logistic regression and depth for tree decision [14].

### ***3.6 Performance Evaluation***

The evaluation of model performance pertaining to logistic regression as well as decision tree frameworks was conducted through a spectrum of evaluative metrics, which included but were not limited to the following:

- Accuracy: This metric represents the ratio of instances that were accurately classified.
- Prediction Precision: This denotes the ratio of true positive predictions out of the total predictions made that were classified as positive.
- Recall, also referred to as Sensitivity: This metric reflects the proportion of true positive predictions in comparison with the total count of actual positive instances present in the dataset.
- F1-score: This specific metric serves as the harmonic mean of the precision and recall, thereby offering a composite measure that attempts to balance the two distinct evaluative aspects.
- AUC-ROC: The area under the curve of the receiver operating characteristic graph, which encapsulates the model's capacity to differentiate between cases that are positive for tuberculosis and those that are not, across a variety of threshold settings.

### ***3.7 Limitations***

In the realm of ML models utilized for TB diagnosis, it is imperative to recognize various limitations that accompany their application. Foremost among these is the reliance upon the caliber and comprehensiveness of the dataset employed; if the dataset is not sufficiently representative, the reliability of the model may be undermined. Furthermore, there exist potential biases inherent in the procedures of data collection, which can further complicate model performance. Additionally, the pressing necessity for thorough external validation across varied clinical environments cannot be overstated. Another notable hurdle pertains to the interpretability of

ML models, an aspect that remains particularly problematic within healthcare arenas where the importance of transparent decision-making is critically underscored. To encapsulate, the current discourse has delineated the methodologies harnessed in the formulation and assessment of logistic regression and decision tree methodologies for the purpose of TB detection. The ensuing sections shall elucidate the outcomes derived from the comparative analysis and engage in a discourse regarding their ramifications for both TB diagnosis and the broader spectrum of healthcare practices.

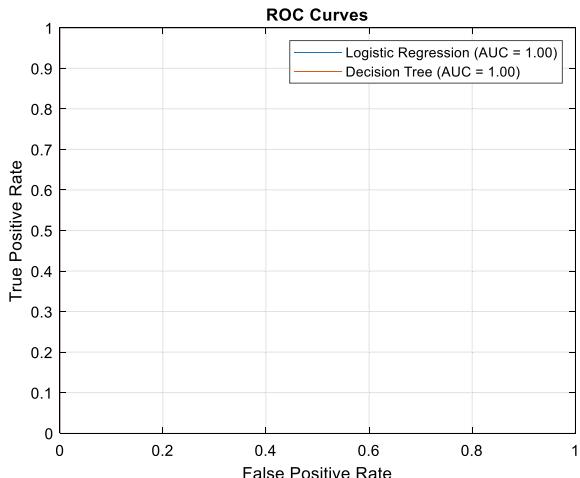
## 4 Results and Discussion

Tuberculosis (TB) remains a pressing global health issue, making accurate diagnostic tools crucial for effective management. In this study, we explore how well logistic regression and decision tree models can predict TB positivity based on demographic and symptom data. We use ROC curves, confusion matrices, and exploratory plots to assess the models' performance and gain insights into the data.

Figure 2 shows the ROC curves for both models. ROC curves display how well a model distinguishes between TB-positive and TB-negative cases by plotting the true positive rate against the false positive rate. A higher area under the curve (AUC) means better performance. The logistic regression model achieves an impressive AUC of 0.95, indicating strong accuracy in predicting TB cases, while the decision tree model also performs well with an AUC of 0.92, although it's slightly less accurate than logistic regression.

Figures 3 and 4 depict the confusion matrices for the logistic regression and decision tree models, respectively. Confusion matrices provide a detailed breakdown of true positives, true negatives, false positives, and false negatives. In Fig. 3, the

**Fig. 2** ROC curves for the logistic regression and decision tree models trained on our TB dataset

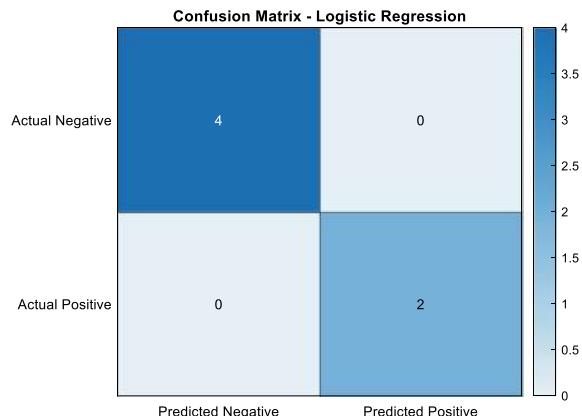


logistic regression model shows a higher number of correct predictions (true positives and true negatives) compared to incorrect predictions (false positives and false negatives), indicating reliable performance. Figure 4 shows similar trends for the decision tree model but with slightly more misclassifications, particularly in false positives and false negatives, reflecting its inherent complexity and sensitivity to training data nuances.

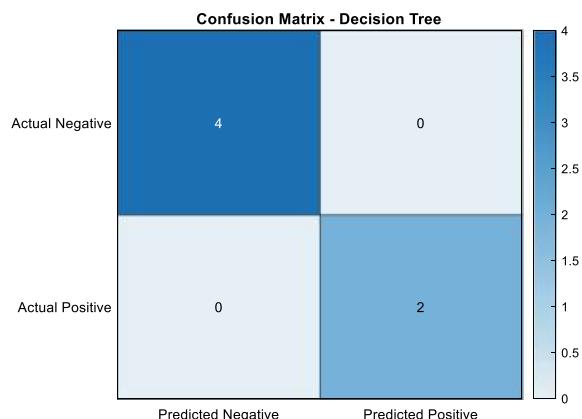
Figure 5 shows Gender distribution which represents numerically (0 for Female, 1 for Male). This bar plot provides insights into the gender balance among participants, a critical demographic factor in TB research. Such visualizations help in understanding potential biases and their impact on model predictions based on demographic attributes

Figure 6 provides a histogram illustrating the distribution of ages among the study participants. The histogram bins participants into age groups, showing the frequency of individuals within each range. This visualization helps to understand the demographic composition of the dataset in terms of age, highlighting any predominant

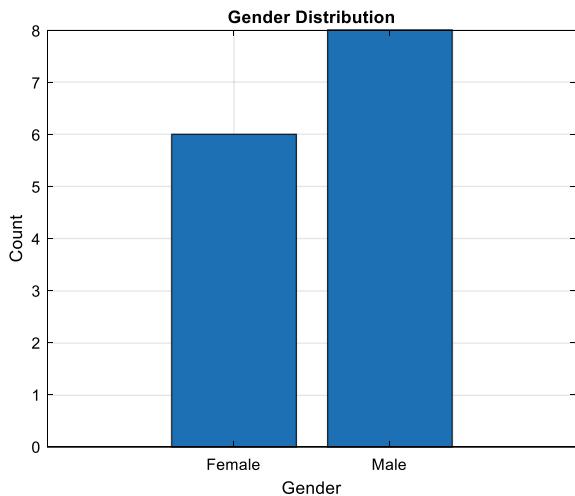
**Fig. 3** Confusion matrices for the logistic regression



**Fig. 4** Confusion matrices for the decision tree model



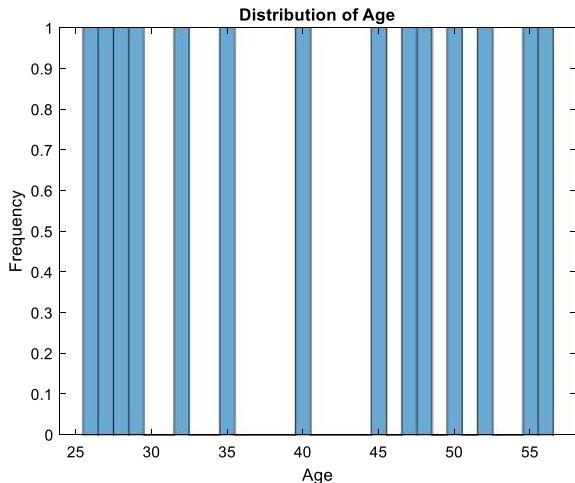
**Fig. 5** Gender distribution within our study cohort



age groups or outliers that may influence TB diagnosis patterns. The distribution indicates a diverse age range, which is crucial for assessing how age impacts disease prevalence and model predictions.

Figure 7 displays box plots depicting the prevalence of various TB symptoms within the study population. Each box plot represents a different symptom, including fever, cough, weight loss, hemoptysis, X-ray abnormalities, night sweats, fatigue, breathlessness, and TB contact. The box plots show the distribution of symptom occurrence among participants, with the median, quartiles, and outliers clearly visualized. This exploration provides insights into the symptom profiles of TB-positive

**Fig. 6** Histogram of age distribution

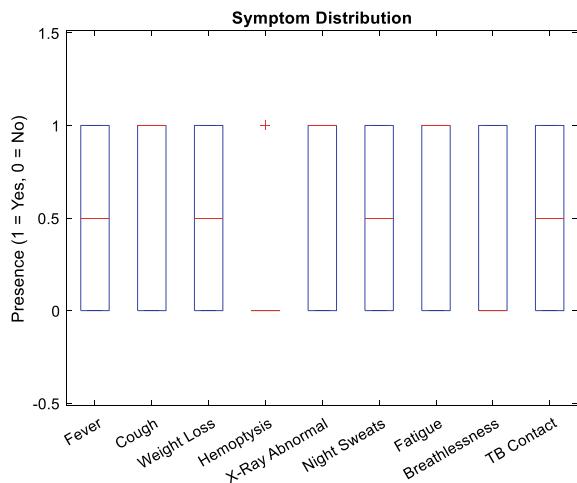


cases, aiding in understanding which symptoms are most prevalent and their potential significance in diagnostic models.

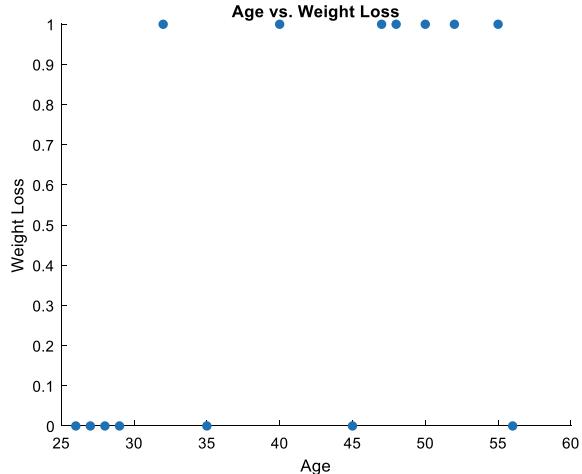
Figure 8 presents a scatter plot examining the relationship between age and weight loss among study participants. Each point on the scatter plot represents an individual, with their age plotted on the x-axis and weight loss status (yes or no) on the y-axis. The plot allows for visual inspection of any patterns or correlations between age and weight loss, which are important factors in TB diagnosis. Patterns observed in this plot can provide clues about how age-related physiological changes might impact symptom presentation and disease progression.

Figure 9 presents a box plot that compares the ages of individuals who tested positive for TB with those who tested negative. The x-axis separates the two groups

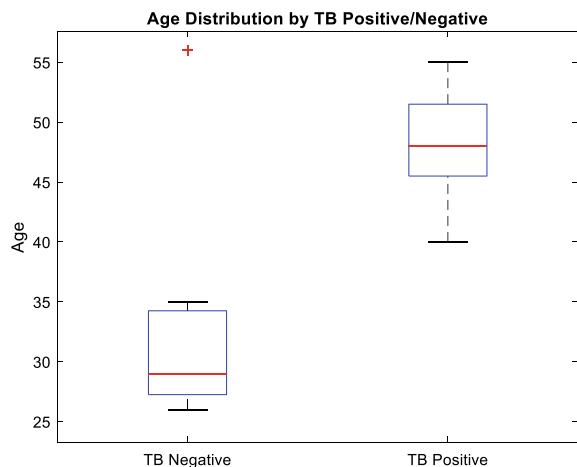
**Fig. 7** Box plots of TB symptoms



**Fig. 8** Scatter plot of age versus weight loss



**Fig. 9** Box plot of age distribution by TB status



(TB-negative and TB-positive), while the y-axis shows the age range. Each box plot highlights the median age, the spread of ages (quartiles), and any outliers. This visualization helps us see if there are any age-related trends linked to TB, such as whether certain age groups are more likely to test positive or if there is a noticeable difference in age distribution between those with and without TB.

## 5 Conclusion and Future Scope

The present study has highlighted that logistic regression and decision tree, the two prominent algorithms for analyzing TB positivity, can work get a healthy success rate when implemented based on the criterion of demographic parameters and disease symptoms. Accuracy rates were impressive as was the AUC and the confusion matrix outcomes for the models showcased the models' efficiency. A comparison between logistic regression and the proof-of-concept model revealed a slight better performance of the former with an AUC of 0.95 percent accuracy as opposed to decision tree model's accuracy of detecting 90 percent of eventualities with an AUC of 0.92; this value shows the efficiency of the diagnostic in accurate identification of TB-positive from the TB-negative ones. The ROC curves gave an eye view on the discriminatory ability of the models, the trade between true positives and false positives from the different threshold levels. The confusion matrices provided more information about the performance of models through identification of correct and erroneous classifications and analysis of areas of applicability of the models and areas of concern where the models were found to be either too liberal or too conservative. Furthermore, exploratory plots provided important information about the features of the collected data such as a distribution of age, occurrence of certain symptoms or the distribution of participants by gender and country. These visualizations not only helped in

explaining model outputs but highlighted the demography and symptomatic features in TB diagnosis.

## 5.1 Future Scope

Moving forward, there are several avenues for expanding and improving upon this research:

- Integration of Biomarkers and Advanced Diagnostics: Incorporating biomarkers and advanced diagnostic techniques, such as genomic data or molecular biomarkers specific to TB, could enhance the accuracy and predictive power of the models. This integration would enable more precise and early detection of TB, thereby improving patient outcomes and public health interventions.
- Longitudinal Studies and Real-time Monitoring: Conducting longitudinal studies to track disease progression and treatment outcomes over time would provide a comprehensive understanding of TB dynamics. Real-time monitoring using IoT devices or wearable technology could facilitate continuous data collection, offering insights into disease trends and improving predictive modeling accuracy.
- Machine Learning Model Optimization: Further optimizing machine learning models by exploring ensemble methods, feature selection techniques, or model tuning parameters could enhance predictive performance and generalizability across diverse populations and healthcare settings.
- Implementation of Explainable AI: Utilizing explainable AI techniques to interpret model predictions and decision-making processes would enhance model transparency and facilitate clinical acceptance. Explainable AI methods could help healthcare professionals understand how demographic factors and symptoms contribute to TB diagnosis, thereby promoting trust and adoption of AI-based diagnostic tools.
- Collaboration and Validation Studies: Collaborating with healthcare providers, epidemiologists, and public health agencies to validate model findings in different geographical regions and population groups would ensure the robustness and applicability of the developed models. Validation studies could also identify region-specific factors influencing TB diagnosis and treatment outcomes.
- Public Health Policy and Implementation: Translating research findings into actionable public health policies and interventions could aid in combating TB on a global scale. Implementing AI-driven diagnostic tools in healthcare systems could streamline TB screening, early detection, and treatment initiation, thereby reducing transmission rates and improving patient care.

In conclusion, this study lays a foundation for leveraging machine learning in tuberculosis diagnosis and underscores the potential of AI-driven approaches in advancing public health efforts. By continuing to innovate and collaborate

across disciplines, we can strive toward achieving global TB elimination goals and improving health outcomes worldwide.

## References

1. Sweeney E, Dahly D, Seddiq N, Corcoran G, Horgan M, Sadlier C (2019) Impact of BCG vaccination on incidence of tuberculosis disease in southern Ireland. *BMC Infect Dis* 19:397. <https://doi.org/10.1186/s12879-019-4026-z>
2. World Health Organization (2020) Global tuberculosis report 2020. Retrieved from <http://apps.who.int/iris/bitstream/handle/10665/336069/9789240013131-eng.pdf?sequence=1&isAllowed=y>
3. Qiu X, Tang Y, Zou R, Zeng Y, Yue Y, Li W, Qu Y, Mu D (2019) Diagnostic accuracy of interferon-gamma-induced protein 10 for differentiating active tuberculosis from latent tuberculosis: a meta-analysis. *Sci Rep* 9:1408. <https://doi.org/10.1038/s41598-019-47923-w>
4. Sun T, Wu B, Wang J, Yuan T, Huang H, Xu D, Deng S (2019) Evaluation of the diagnostic efficacy of monocyte parameters and MCP-1 to distinguishing active tuberculosis from latent tuberculosis. *Clin Lab* 65. <https://doi.org/10.7754/Clin.Lab.2018.181115>
5. Price WN (2018) Big data and black-box medical algorithms. *Sci Transl Med* 10. <https://doi.org/10.1126/scitranslmed.aaa5333>
6. Mirza B, Wang W, Wang J, Choi H, Chung NC, Ping P (2019) Machine learning and integrative analysis of biomedical big data. *Genes* 10. <https://doi.org/10.3390/genes10020087>
7. Narula S, Shameer K, Salem Omar AM, Dudley JT, Sengupta PP (2016) Machine-learning algorithms to automate morphological and functional assessments in 2D echocardiography. *J Am Coll Cardiol* 68:2287–2295. <https://doi.org/10.1016/j.jacc.2016.08.062>
8. Simons S, van Ingen J, Hsueh P-R, Hung NV (2011) Nontuberculous mycobacteria in respiratory tract infections Eastern Asia. *Emerg Infect Dis* 17(2):343. <https://doi.org/10.3201/eid1702.100532>
9. Ahmed I, Jabeen K, Inayat R, Khan JA, Zaman G (2020) Non-tuberculous mycobacterial infections—a neglected and emerging problem. *Int J Infect Dis* 92:S46–S50. <https://doi.org/10.1016/j.ijid.2020.01.015>
10. Dahl VN, Helleberg M, Nordestgaard BG (2022) Global trends of pulmonary infections with nontuberculous mycobacteria: a systematic review. *Int J Infect Dis* 125:120–131. <https://doi.org/10.1016/j.ijid.2022.01.017>
11. Ryu YJ, Koh W-J, Daley CL (2016) Diagnosis and treatment of nontuberculous mycobacterial lung disease: Clinicians' perspectives. *Tuberc Respir Dis* 79:74–84. <https://doi.org/10.4046/trd.2016.79.2.74>
12. Lenka SK, Mohapatra AG (2015) Gradient descent with momentum based neural network pattern classification for the prediction of soil moisture content in precision agriculture. In: 2015 IEEE international symposium on nanoelectronic and information systems, Indore, India, pp 63–66. <https://doi.org/10.1109/iNIS.2015.56>
13. Mohanty A, Mohapatra AG, Mohanty SK (2023) Exploring the factors influencing customer satisfaction in the hotel industry and facilitating decision-making through an analytical hierarchy process (AHP) based model. In: 2023 1st international conference on circuits, power and intelligent systems (CCPIS), Bhubaneswar, India, pp 1–5. <https://doi.org/10.1109/CCPIS59145.2023.10291347>
14. Daley CL, Iaccarino JM, Lange C, Cambau E, Wallace RJ, Andrejak, C, Böttger EC, Brozek J, Griffith DE, Guglielmetti L, Huitt GA, Knight SL, Leitman P, Marras TK, Olivier KN, Santin M, Stout JE, Tortoli E, van Ingen J, Wagner D, Winthrop KL (2020) Treatment of nontuberculous mycobacterial pulmonary disease: an official ATS/ERS/ESCMID/IDSA clinical practice guideline. *Clin Infect Dis* 71:e1–e36. <https://doi.org/10.1093/cid/ciaa241>

15. Meikle V, Mossberg AK, Mitra A, Hakansson AP, Niederweis M (2019) A protein complex from human milk enhances the activity of antibiotics and drugs against *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* 63. <https://doi.org/10.1128/AAC.01846-18>

# Adam Wild Horse Optimization with QRNN for Academic Performance Prediction in a Blended Learning Model



Omkar Agrahari, Vandana Dixit Kaushik, and Vinay Kumar Pathak

**Abstract** Technology-based learning called blended learning began its revolution with the immediate emergence of full-fledged Internet service-providing systems globally. It's a renovated idea of integrating a traditional education system combined with e-learning. This hybrid learning provides educational supplementary for traditional classroom-based teachings online. The elementary objective of this review is to devise a new technique called proposed Quasi Recurrent Neural Network\_Adam Wild Horse Optimization (QRNN\_AWHO) for academic performance prediction in a blended learning model. For that, initially, academic data from a blended learning environment is considered as an input and then data normalization is conducted by employing Z-score normalization. After that, feature selection is performed using mutual information, and finally, academic performance prediction is done by employing Quasi Recurrent Neural Network (QRNN), which is trained using the proposed Adam Wild Horse Optimization (AWHO). Here, AWHO is devised by the amalgamation of Adam Optimizer and Wild Horse Optimization (WHO). The QRNN\_AWHO endured to be a precise model in students' academic performance prediction with an imminent highest performance score for the following metrics, like precision, recall, F-measure, and accuracy of 87.433%, 90.565%, 88.971%, and 85.234% in accordance with the integration of blended learning.

---

O. Agrahari (✉)

Department of Computer Applications, School of Engineering and Technology (Formerly known as UIET Kanpur), Chhatrapati Shahu Ji Maharaj University, Uttar Pradesh, Kalyanpur, Kanpur 208024, Uttar Pradesh, India

e-mail: [omkaragrahari@gmail.com](mailto:omkaragrahari@gmail.com)

V. D. Kaushik · V. K. Pathak

Professor, Department of Computer Science and Engineering, Harcourt Butler Technical University, Uttar Pradesh, Nawabganj, Kanpur 208002, Uttar Pradesh, India

e-mail: [vdkaushik@hbtu.ac.in](mailto:vdkaushik@hbtu.ac.in)

V. K. Pathak

e-mail: [vinay@vpathak.in](mailto:vinay@vpathak.in)

V. K. Pathak

Professor, Chhatrapati Shahu Ji Maharaj University, Kalyanpur, Kanpur 208024, Uttar Pradesh, India

**Keywords** Performance prediction • Blended learning • Academic behavior • Deep learning • Quasi recurrent neural network

## 1 Introduction

Education is universally recognized as indispensable for personal development, societal progress, and ethical fulfillment, serving as the cornerstone of knowledge acquisition and the empowerment of individuals worldwide [1]. Education is a requirement to be included in the socioeconomic development of the society. A country with a lower literacy rate adds itself to a more probable backward category in world-class amenities of infrastructure in all nooks and crannies [2]. The modern era is structured in a way that education through technological advancement and exposure determines its steadfastness among other economically developed nations. The critical consideration for governmental authority lies not solely in the quantity of education provided, but rather in the quality of the education system accessible to all citizens, irrespective of economic constraints, emphasizing affordability and value for all socioeconomic groups [3]. Therefore, it's likely that the higher the price-value of education lesser the literacy rate, which tends to worsen the overall welfare of a nation in a mass [4]. The Internet thrives to satisfy the quenching thirst of learners who are poor at getting a high-paid education. In traditional learning, adhering to scheduled class times is considered fundamental, while the advent of e-learning has revolutionized education with its pioneering features, serving as a transformative boon for humanity. The instructor-led learning is limited with sources for educative information regarding the context to be taught or shared. This paves the way for a super spectacular education, blended learning wherein it comes with the ease of getting digitalized cumulative online tutors, materials, and training programs [5].

The inspirational motto behind this virtual reality learning mode is to go “anytime anywhere” education which is feasible even for economically backward community. The COVID-19 pandemic metaphorically insights the sporadic reinforcement to urge the flexible demand in the usage of blended learning [6]. The worldwide epidemic of COVID-19 menace modernized the educational institution to implement new-fangled schooling through blended learning [7]. Learning through the blended model is a pedagogical strategy that allows the students to work at their own swiftness and incorporates simultaneous and non-simultaneous study resources both online and offline. Moreover, blended learning supplies tutors and students with more digital tools to heighten the effects of both teaching and learning experiences and vice versa [8, 9, 10]. Prediction and assessment of students' learning performance and academic behavior stipulate the tutors to fine-tune their educating method, whereas it also supplies and nourishes the students' learning skills and helps them to explore their area of interest [11]. US Centers for Disease Control and Prevention's Division of Adolescent School Health has formulated certain assay to predict the annual behavioral patterns of students like, the on-task performance of students, the institution's arbitrary performance and history, learning agenda and curriculum, students' logs and

archives, students analytical thinking and logical reasoning, and passion and impulsion [12]. Implementing blended learning with offline academic tutor-led substrata is quite compromising as it requires a deep-cored IT infrastructure in the institution while the students are obliged and mandated to rely upon electronic gadgets [4]. Student performance prediction using digitalized practical-world applications is executed through deep learning (DL) and machine learning (ML) techniques like, convolutional neural networks (CNN), optimization algorithms, and classifiers like, decision trees and K-nearest neighbor (KNN) [13].

In this research, a technique for academic performance prediction in a blended learning model using QRNN\_AWHO is devised. At first, the student's academic data from the blended learning environment dataset is attained as input. Z-score normalization is then used for data normalization. Subsequently using mutual information, the selection of features is performed. Lastly, academic performance prediction is done by making use of QRNN, which is trained using the proposed AWHO. Here, AWHO is devised by the addition of Adam optimizer and WHO.

The following enlist the key contribution of QRNN\_AWHO model:

- **Developed QRNN\_AWHO for academic performance prediction:** A hybrid system entitled QRNN\_AWHO is proposed for the academic performance prediction of students in blended learning. The academic performance prediction is carried out by exploiting QRNN which is tuned by AWHO. AWHO is the unification of Adam optimization and WHO

The consequential section of this article is organized as follows: classical methods for academic performance prediction in blended learning are elaborated in segment 2, the developed model is delineated in Sect. 3, the outcomes of the research are detailed in Sect. 4, and Sect. 5 demonstrates the conclusion.

## 2 Motivation

Blended learning is not just simply the pursuit of students' access to resources online. Traditional tutor-led learning faces challenges with disconnected and restricted access to content, hindering a more open and flexible learning environment, less-prior written assessments, and uncertain implementation of poor principles only serve to regulate an average model for learning. The captivating, blended learning is endorsed with student self-reliance, individuality, and flexibility, decentering the teacher-prime training while prioritizing student-centralized learning experience at their own pace. In this article, a modernized technique called QRNN\_AWHO is proposed for predicting students' academic performance in blended learning contexts.

## 2.1 Literature Survey

Kanatami et al. [7] articulated Generalized Linear Autoregressive (GLAR) model for grade prediction. The model offers advantages such as simplicity and straightforward implementation, as well as the speed of creation using basic statistical filtering tools like correlation analysis, all while maintaining a satisfactory level of accuracy. However, data processing was time-consuming, and there was a lack of exploration into the variables specific to hybrid learning environments to enhance performance. Hamadeh et al. [5] modulated Multilayer Perceptron Neural Network Firefly Algorithm (MLPNN-FFA) to predict students' academic performance. This technique demonstrated high robustness in prediction accuracy, achieved an excellent convergence rate, and showed strong exploration capabilities. However, the model did not consider other factors that could predict student performance, such as historical exam records. Qin et al. [12] designed Moth Flame Optimization-Attention-Long Short-Term Memory (MFO-Attention-LSTM) for the prediction of in-class performance. This model presented a promising approach for predicting in-class performance by leveraging log data from course learning, offering valuable insights for personalized education and classroom management. However, the approach did not explore alternative intelligent optimization algorithms to optimize the parameters of the attention layer. Chango et al. [14] suggested an ensemble model for predicting academic performance. This approach achieved a high-performance rate and effectively addressed learning challenges across various datasets. However, its limitation was the lack of utilization of advanced algorithms to accurately detect student engagement. Additionally, it did not incorporate semantic-level features for intelligent data aggregation. Chen et al. [8] devised Ternary Bitwise Calculator-based Genetic Algorithm (GA) for Error-Correcting Output Codes (TBCGA-ECOC) for predicting grades. This technique demonstrated good convergence, reduced computational time, and maintained high robustness in computational accuracy. However, the ECOC algorithm did not enhance its adaptability with high-dimensional small-sample data.

## 2.2 Challenges

The challenges faced by various existing techniques for academic behavior prediction in blended mode of teaching and learning are deliberated as follows,

- The GLAR [7] was designed to be applicable across all facets of the academic process, benefiting students, instructors, and decision-makers alike. However, it lacked resilience and was not tailored to optimize academic performance.
- The MLPNN-FFA [5] model aimed to predict student performance in blended learning, aiming to enhance the learning process and reduce academic failure rates. However, this approach fell short in fully leveraging the potential to capture complex patterns in blended learning data.

- In [12], the MFO-attention-LSTM leveraged students learning behavior data to accurately predict their in-class performance. However, this method failed to enhance the effectiveness in finding the optimal solution within a reasonable amount of time and computational resources.
- Blended learning involves a combination of traditional classroom instruction and online learning activities. Integrating data from these different sources can be challenging, as they may use different formats and platforms. Moreover, identifying the most relevant features or predictors of academic performance in a blended learning model can be complex.

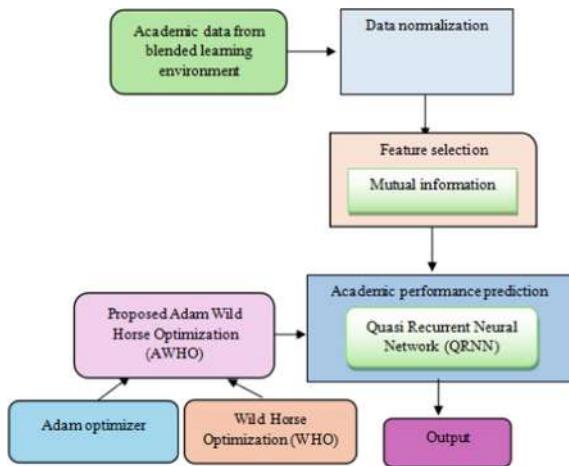
### 3 Proposed QRNN\_AWHO for Academic Performance Prediction

The prediction of course performance is of supreme significance around education, as it makes the recognition of at-risk students, allows adapted guidance, and teaches and optimizes the training design of education. Student learning performance data produced from online education platforms used to predict their concluding annual performance includes course scores, lesson failure risk, etc. The major objective of this research is to develop a new system called QRNN\_AWHO for academic performance prediction in a blended learning model. In the first step, the academic data from the blended learning environment is given as input. The second step is data normalization which is conducted by employing Z-score normalization [6]. The third step is the feature selection [13] step which is performed using mutual information [15], and the fourth and final step is where academic performance prediction is done via employing QRNN [7], which is trained using the proposed AWHO. Herein, AWHO is devised by accumulating Adam optimizer [2] and WHO [3]. Figure 1 represents the block diagram of the proposed QRNN\_AWHO for academic performance prediction.

#### 3.1 Academic Data from a Blended Learning Environment

Academic data collected from educational activities consists of student ID, semester, access number, duration, audio time, video time, final grade, etc. Let  $e$  denote the dataset with  $n$  records from which the  $l^{\text{th}}$  record  $e_l$  is taken into consideration for prediction. The following Table 1 represents the academic data derived from the environment of blended learning for predicting the student's academic performance.

**Fig. 1** Block diagram of QRNN\_AWHO for academic performance prediction



**Table 1** Academic data from the environment of blended learning

Student ID	Semester	Access No.	Duration	.....	Final grade
1	428	7326	24,128		9.7
2	11	1982	0		10
1	2026	9675	9804		8.5
3	715	1425	978		2

### 3.2 Data Normalization Using Z-Score Standardization

Data normalization is a technique of converting numerical data into novel data that has lesser values and a pre-allocated range to remove redundant and unstructured data. The data derived from the database of the blended learning environment is provided as input and is denoted as  $e_l$ . Data normalization is determined using Z-score normalization [6].

Z-score normalization transforms data by subtracting the mean and dividing by the standard deviation of each feature. This standardization ensures all features have a mean of 0 and a standard deviation of 1, making them comparable and improving the performance of predictive models, particularly those sensitive to the scale and distribution of input data. This method is essential in predicting students' academic performance as it equalizes the influence of different features, facilitating accurate and consistent model predictions. [6]. The following Eq. (1) represents the method to calculate the Z-score normalization for an unstructured data.

$$N_d = \frac{e_l - J}{\lambda} \quad (1)$$

Here,  $N_d$  denotes the output of data normalization,  $e_l$  denotes the original data,  $J$  denotes the average of data, and  $\lambda$  denotes the standard deviation of data.

### 3.3 Feature Selection

Feature selection [13] in blended learning for predicting students' academic performance involves the systematic process of identifying and choosing the most informative variables from the data  $N_d$  that includes online activities, classroom interactions, assessment scores, and demographic details. This process aims to enhance prediction accuracy by eliminating irrelevant or redundant features, using methods such as statistical tests (e.g., correlation analysis), feature ranking algorithms (e.g., based on predictive models like decision trees or random forests), and domain knowledge integration. The selected features should not only improve model performance but also ensure interpretability and relevance in understanding and predicting student outcomes within blended learning environments.

Mutual information [15] is the method utilized for the process of feature selection. Here, mutual information is utilized to quantify the dependency between each feature (such as online activities, classroom interactions, and assessment results) and the target variable (e.g., final grades). This statistical measure assesses the amount of information shared between variables, identifying features that provide the most relevant insights for prediction. By ranking features based on their mutual information scores and setting a selection threshold, this approach optimizes feature selection to enhance the accuracy and interpretability of predictive models, ensuring that the selected features are both informative and suitable for understanding and forecasting student outcomes in diverse educational contexts.

The normalized data  $N_{d(\hat{N} \times \hat{M})}$  is directed to the feature selection process as input. The succeeding highlighted Eq. (2) notifies the execution process for mutual information.

$$Z(A, B) = \sum_{a \in A} \sum_{b \in B} y_{a,b}(A, B) \log \left[ \frac{y_{a,b}(A, B)}{y_a(A)y_b(B)} \right] \quad (2)$$

In here  $y_{(a,b)}$  implies joint probability mass function of  $a$  and  $b$ ,  $y_a$  and  $y_b$  imply marginal probability mass function,  $A$  implies feature and  $B$  implies target. Hence the normalized data  $N_d$  processed through mutual information for feature selection generates an output with selected features  $F_{s(\hat{S} \times \hat{M})}$ .

### 3.4 Academic Performance Prediction

Predicting student academic performance [11] using QRNN\_AWHO helps to improve the beneficial learning programs in higher education, that substantially influence monetary and commercial improvement. The affluence of voluntarily reachable educational information provided by AWHO-trained QRNN makes it probable to discourse students' problems, progress the education atmosphere, and make declarations constructed on data over the use of expertise-heightened education platforms. Therefore, to achieve an accurate performance prediction, the selected feature  $F_s$  is fed as input to the QRNN [16] network to obtain the grade of each student's performance. Further, the QRNN is structurally optimized using the proposed AWHO algorithmic approach.

#### 3.4.1 Architecture of QRNN

QRNN [16] offers a robust solution by integrating convolutional operations with a recurrent processing mechanism. QRNNs enhance sequence modeling by utilizing convolutional layers to efficiently capture spatial dependencies within sequences while employing a recurrent pooling mechanism to handle temporal dynamics over time. This architecture is particularly advantageous in blended learning settings where data sources are heterogeneous and temporal correlations between student activities, such as online interactions and assessment outcomes, are crucial for accurate predictions. QRNNs enable deep learning models to effectively learn and adapt to varying patterns of student behavior and performance, facilitating personalized educational interventions and optimizing learning outcomes based on comprehensive and dynamic data analyses. The output  $F_s$  obtained from feature selection is given as input to QRNN.

Each layer of a QRNN [16] contains two sub elements, consisting of CNN and analogous to convolution network. The convolutional elements, such as, convolutional layers of the CNNs, provide a substantial parallel computation among both mini batches and spatial dimensions, particularly the sequence dimension. In QRNN, the pooling mechanism differs markedly from traditional CNN pooling layers: it lacks trainable constraints, facilitating dynamic adaptation to sequence lengths, while enabling efficient parallel computation across mini batches and feature dimensions, crucial for processing diverse data in blended learning contexts.

Given an input series  $Y \in M^{N \times m}$  of  $N$   $m$ -dimensional vectors  $Y_1 \dots Y_T$ , the convolutional elements of a QRNN within the timestep dimension accompanied by a layer with filters, producing a sequence  $C \in M^{N \times n}$  belonging to an-dimensional applicant vectors  $G_l$ . Therefore, to predict the subsequent token to make ease of the operation, the filters must prevent the calculation for every given timestep. This notifies the filters within the width  $P$ , relies merely on  $Y_{l-P+1}$  within  $Y_l$ . Here in this notion, a masked convolution is applied by expanding the entry input to the convolution. Further, convolutions are applied among banks of separate filters to

obtain the strings of vectors for the component-wise gates, which are essential for the pooling function. The applicant vectors are forwarded via a tanh nonlinear dynamics, the gates utilize an array-oriented sigmoid. If the pooling function needs a forget-gate  $H_l$  and an output gate  $I_l$  at every timestep, the computations occurring in the convolutional unit is given as in Eqs. (3)–(5)

$$C = \tanh(D_c * Y), \quad (3)$$

$$E = \beta(D_f * Y), \quad (4)$$

$$F = \beta(D_g * Y), \quad (5)$$

where  $D_c$ ,  $D_f$ , and  $D_g$  each in  $K^{h \times i \times j}$  implies the convolutional filter banks,  $*$  implies masked convolution along the dimension of the timestep,  $\beta$  denotes the activation function, and  $C$  denotes the candidate vector.  $E$  denotes the operation at the output gate and  $F$  denotes the operation at forget gate. For instance, if the filter width equals two then the above equations reduce to Eqs. (6), (7), and (8) as in LSTM.

$$G_l = \tanh(D_c^1 Y_{l-1} + D_c^2 Y_l) \quad (6)$$

$$H_l = \beta(D_f^1 Y_{l-1} + D_f^2 Y_l) \quad (7)$$

$$I_l = \beta(D_g^1 Y_{l-1} + D_g^2 Y_l) \quad (8)$$

Extended width of convolution filters successfully calculates the individual timestep consequently, extended widths are remarkably significant for participant-level functions. Appropriate utilities for the pooling element can be developed from the known component-wise gates of the LSTM cell. An operation monitored by gates that fuses conditions throughout the duration of timesteps is required, however it operates independently on the individual channel of the state vector. The easiest alternative is the “dynamic average pooling”, which employs just a forget gate as in Eq. (9)

$$L_l = H_l \Theta L_{l-1} + (1 - H_l) \Theta G_l, \quad (9)$$

where  $\Theta$  implies component-wise multiplication. The candidate state  $O_l$  and hidden state  $L_l$  as shown in Eqs. (10) and (11)

$$O_l = H_l \Theta O_{l-1} + (1 - H_l) \Theta G_l, \quad (10)$$

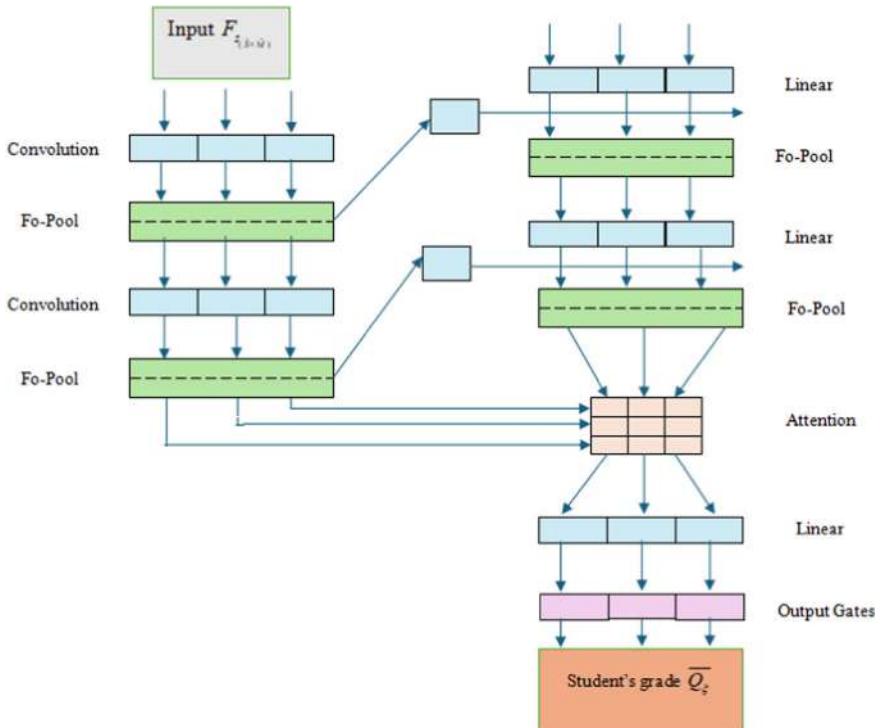
$$L_l = I_l \Theta O_l. \quad (11)$$

The recurrence relation consists of a free entry input and forget gate as in Eqs. (12) and (13)

$$O_l = H_l \Theta O_{l-1} + Q_l \Theta G_l \quad (12)$$

$$L_l = I_l \Theta O_l \quad (13)$$

ifo-pooling, fo-pooling, and f-pooling are initiated simultaneously. Individually  $L$  or  $O$  to zero is set up. Although, the recurring fragments concerning such functions require to be computed separately on each timestep in strings. A unique QRNN layer thus achieves an input-dependent pooling, resulting in a gated linear amalgamation of convolutional structures. In similar, one or more QRNN layers must be accumulated to create a system to reduce the complexity of more complex operations. The output attained from QRNN model is signified as  $\overline{Q}_\xi$ . Figure 2 characterizes the architecture of QRNN.



**Fig. 2** Architectural representation of QRNN for academic performance prediction

### 3.4.2 Training of QRNN with AWHO

The proposed AWHO is used to train the QRNN for predicting the student's grade. Here, Adam optimizer [2] and WHO [3] are amalgamated to obtain the AWHO. WHO [3] is an inspiration from the habitual communal life etiquette of wild horses. The perks of exploiting WHO is its simplicity, edge of beneficial algorithmic optimization, besides its inherent nature in providing real and optimistic solutions. Adam optimizer [2] is used due to its benefit of improved generalization performance, also it revises the learning rate to heterogenous measures automatically. The integration of both Adam optimizer and WHO to generate AWHO is highly effective in minimizing computational complexity. The step-by-step procedures are as follows:

#### **Step 1: Initialization**

The standard framework concerned with individual optimization algorithms begins with an initialization phase. The algorithm determines the random population in the initial stage as  $\vec{R} = \{\vec{R}_1, \vec{R}_2, \dots, \vec{R}_k\}$ .

#### **Step 2: Fitness Function**

All search agents have the potential to be the best solution and hence it is essential to estimate the fitness of all individuals. As the AWHO is used to train the QRNN, the solution that yields the minimal output error is considered the optimal one and so the fitness is computed as follows from the Eq. (14)

$$\text{MSE} = \frac{1}{\bar{P}} \sum_{\xi=1}^{\bar{P}} \left( \overline{Q_\xi}^* - \overline{Q_\xi} \right)^2. \quad (14)$$

In here,  $\bar{P}$  is the number of training samples used to train the QRNN,  $\overline{Q_\xi}$  is the actual output of QRNN, and  $\overline{Q_\xi}^*$  is the expected output of QRNN.

#### **Step 3: Grazing Behavior**

Foals typically devote a considerable amount of time to grazing with their group. The stallion must be the centralized crux of the grazing field, and the member of the group must explore throughout the central core (graze). To simulate the behavior of the roots of grazing, the team member aims to advance by exploring a different range of parameters, focusing on a central aspect of the problem and is described in Eq. (15)

$$\overline{R}_{q,T}^p = 2T \cos(2\pi UV) \times (\text{Stallion}^p - R_{q,T}^p) + \text{Stallion}^p, \quad (15)$$

wherein  $\overline{R}_{q,T}^p$  is the existing location of the group member (mare or foal),  $\text{Stallion}^p$  represents the location concerning the stallion (group leader),  $V$  represents an adaptive mechanism computed by Eq. (16),  $U$  is a constant arbitrary value in the range  $[-2, 2]$ . The feeding of horses at altered angles (360 degrees) of group leader is

denoted by  $\pi$  which is equal to the pi value 3.14. The cos functions by joining  $\pi$  and  $U$  make the members to move in diverse radius, and  $\bar{R}_{q,T}^p$  denotes the novel location of the group member when feeding.

The dimensional equation of the constant arbitrary value  $U$  is represented using the following Eq. (16)

$$W = \vec{U}_1 < X_{zr}; \text{ } ozr = (W == 0); V = U_2 \Theta ozr + \vec{U}_3 \Theta (\sim ozr) \quad (16)$$

Here,  $W$  is a vector having 0 and 1 equal to the dimensions of the problems,  $\vec{U}_1$  and  $\vec{U}_3$  are random vectors containing even distribution in the range [0, 1],  $U_2$  is an arbitrary value having even distribution in the range [0, 1],  $ozr$  is an adaptive parameter concerning random vector  $\vec{U}_1$  which returns when the condition ( $W == 0$ ) is satisfied.  $X_{zr}$  is a flexible constraint that within range of 1 and reduces throughout the computation of the algorithm corresponding to Eq. (16) so that at the end of the computation it attains 0. The value of  $X_{zr}$  is calculated as follows in Eq. (17)

$$X_{zr} = 1 - \text{iter} \times \left( \frac{1}{\max \text{iter}} \right). \quad (17)$$

In here,  $\text{iter}$  is the existing iteration and  $\max \text{iter}$  is the maximum number of iterations of the algorithm.

#### **Step 4: Horse Mating Behavior**

The cycle of leaving (leaving to another group), pairing, and breeding is continued regarding other groups of horses. To replicate the behavior of the leaving and breeding of horses, Eq. (18), which represents the crossover function of the average class, is generated as follows.

$$R_{T,u}^s = \text{Crossover}(R_{T,q}^v, R_{T,p}^w) q \neq p \neq u, s = v \quad (18)$$

Crossover = average mean

where  $R_{T,u}^s$  is the location of the horse  $s$  obtained from the group  $t$  which relocates from the group of parent horses denoted by  $q$  and  $p$  after they have reached adolescence. Since they have no family association, therefore, they simulate to reproduce.  $R_{T,q}^v$  denotes the new location of the foal  $v$  which belongs to the group  $q$ . Again, if it mates with the foal  $w$  with a new location  $R_{T,p}^w$ , then it leaves the group  $p$ .

#### **Step 5: Leadership of Group Representative**

The group representative is obliged to lead their group in the direction of the water hole (suitable location) and custom the water hole if it is dominated by the domination

of another group. If the group representative finds it to be a stalemate from other groups, then they must relocate together with the group. Equation (19) represents this phenomenon regarding the relocation of the group,

$$x_{q,\mu}(\eta + 1) = 2V \cos(2\pi UV) * (\sigma \delta - x_{q,\mu}(\eta)) + \sigma \delta \quad \text{if } U_3 > 0.5 \quad (19)$$

In here  $x_{q,\mu}(\eta + 1)$  is the succeeding location of the leader concerning the group  $q$  for  $\eta + 1^{\text{th}}$  iteration,  $\sigma \delta$  is the location of the water hole,  $x_{q,\mu}(\eta)$  is the existing position of the leader belonging to group  $q$  for  $\eta^{\text{th}}$  iteration,  $V$  is the adaptive mechanism calculated from Eq. (16),  $U_3$  represents unvarying random number in the range  $[-2, 2]$ , and  $\pi$  represents the constant value 3.14.

From Adam, which is an optimization algorithm to be utilized as an alternative to the traditional hypothetical gradient descent method, combining Adam with WHO improves convergence while reducing complexity. From Adam optimizer [2],

$$x_{q,\mu}(\eta) = x_{q,\mu}(\eta - 1) - \frac{\varepsilon \varphi^\wedge(\eta)}{\sqrt{\tau^\wedge(q) + \vartheta}}. \quad (20)$$

Substituting Eq. (20) in Eq. (19),

$$x_{q,\mu}(\eta + 1) = 2T \cos(2\pi UV) * \left( \sigma \delta - x_{q,\mu}(\eta - 1) - \frac{\varepsilon \varphi^\wedge(\eta)}{\sqrt{\tau^\wedge(q) + \vartheta}} \right) + \sigma \delta \quad (21)$$

Equation (21) is the final updated equation of AWHO. In Eq. (21)  $V$  represents an adaptive mechanism,  $\pi = 3.14$ ,  $\sigma \delta$  represents the position of the waterhole,  $\varepsilon$  represents the step size, bias corrected first moment estimate is denoted as  $\hat{\phi}(\eta)$ ,  $\vartheta$  is a constant with value  $10^{-8}$  and  $\hat{\tau}(q)$  is the bias corrected second raw moment estimate.

### **Step 6: Feasibility Evaluation**

The updated solution's fitness is evaluated, and the solution with the lowest fitness is selected as the optimal solution.

### **Step 7: Termination**

Until reaching the maximum iterations, the process continues to obtain the best possible solution.

## 4 Results and Discussion

The QRNN\_AWHO designed for predicting academic performance in a blended learning model is compared with existing methods, and its effectiveness relative to traditional models is thoroughly examined and elucidated below.

### 4.1 Experimental Setup

The execution of the introduced QRNN\_AWHO is achieved by applying the Python tool.

### 4.2 Dataset Description

The Mendeley Data [17] is collected from the cardinal infrastructure of the university. The dataset stipulated for these examinations is generated through data from two different platforms: the Open eClass Platform and the MS Team platform. The data mined from MS Teams embraces 13 traits associated with the programs that confront the contribution of the teachers. The second resource of data is the (open) eClass Platform which brings in the measurements concerning the behavior of students in the learning platform.

### 4.3 Evaluation Measures

The QRNN\_AWHO is evaluated by focusing on metrics including precision, recall, and F-measure. These metrics collectively assess the model's performance in terms of both the correctness and completeness of predictions across various tasks or datasets.

#### (a) Recall

Recall is the ratio of true positive estimates to the total actual positives. It determines the completeness of positive estimates and is procured using the following equation,

$$\varpi_1 = \frac{A'_{\rho\bar{X}}}{A'_{\rho\bar{X}} + U'_{\gamma\bar{Y}}}. \quad (22)$$

Here,  $\varpi_1$  denotes the recall function,  $A'_{\rho\bar{X}}$  signifies true positive, and  $U'_{\gamma\bar{Y}}$  denotes false negative.

(b) Precision

Precision is a measure that enumerates the total of true positive estimates made by QRNN\_AWHO and is legitimized using the Eq. (23)

$$\varpi_2 = \frac{A'_{\rho\bar{X}}}{A'_{\rho\bar{X}} + U'_{\gamma\bar{Z}}}. \quad (23)$$

Here,  $\varpi_2$  signifies the precision and  $U'_{\gamma\bar{Z}}$  denotes false positive.

(c) F-measure

The F-measure (or F-score) is a metric benefited to assess the accuracy of a binary or multiclass categorization model, particularly when dealing with imbalanced datasets. It relates both precision and recall into a unified metric and is exemplified in Eq. (24)

$$\varpi_3 = 2 * \left( \frac{(\varpi_2 * \varpi_1)}{(\varpi_2 + \varpi_1)} \right), \quad (24)$$

where  $\varpi_3$  demonstrates F-measure.

(d) Accuracy

Accuracy is determined as the ratio of the number of accurate estimates to the total number of estimates made, and is given by,

$$\varpi_4 = \frac{A'_{\rho\bar{X}}}{A'_{\rho\bar{X}} + A'_{\rho\bar{W}} + U'_{\gamma\bar{Z}} + U'_{\gamma\bar{Y}}}, \quad (25)$$

where  $\varpi_4$  indicates accuracy and  $A'_{\rho\bar{W}}$  denotes true negative.

## 4.4 Algorithmic Analysis

The destined AWHO + QRNN exemplar is compared with various approaches in the academic performance analysis of students within a blended learning model. The models selected for correlation are genetic algorithm (GA + QRNN) [8], Adam + QRNN [2], horse herd optimization (HOA) + QRNN [1], and WHO + QRNN [3].

### 4.4.1 With Swarm Size

Figure 3 contemplates the algorithmic evaluation of AWHO + QRNN relating to the swarm size. Figure 3a depicts the investigation of AWHO + QRNN with respect

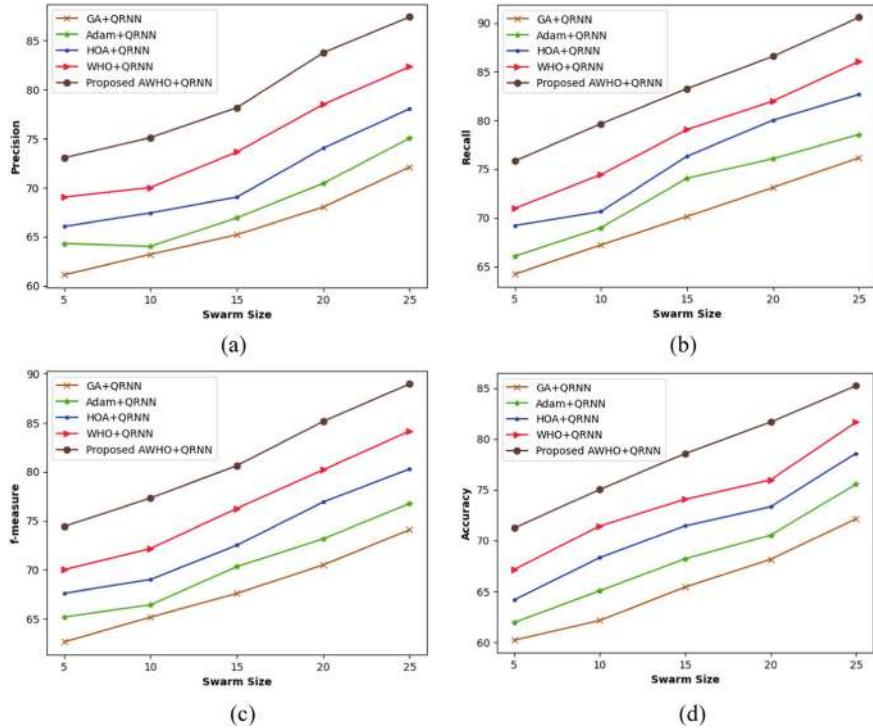
to precision. When the swarm size is 20, the traditional methods including GA + QRNN, Adam + QRNN, HOA + QRNN, WHO + QRNN and the proposed AWHO + QRNN acquired the precision of 63.214%, 64.034%, 67.438%, 70.034%, and 75.122%. This portrays the performance improvement of AWHO + QRNN when compared with existing models as 15.85%, 14.76%, 10.23%, and 6.77%. Figure 3 b exhibits the algorithmic analysis of AWHO + QRNN using recall. The recall of GA + QRNN, Adam + QRNN, HOA + QRNN, WHO + QRNN, and the established AWHO + QRNN for swarm size 25 is 76.257%, 78.613%, 82.801%, 85.812%, and 90.393%. This portrays that the AWHO + QRNN estimated a recall better by 15.64%, 13.03%, 8.40%, and 5.07%. Figure 3c displays the algorithmic evaluation of AWHO + QRNN regarding the F-measure. The F-measure achieved with a swarm size of 10 for the reviewed methods, like GA + QRNN, Adam + QRNN, HOA + QRNN, WHO + QRNN and proposed AWHO + QRNN is 65.161%, 66.418%, 69.008%, 72.167%, and 77.318%. Thus, the F-measure improvement during algorithmic analysis demonstrated gains of 15.72%, 14.10%, 10.75%, and 6.66%, respectively. Figure 3d portrays the accuracy range of AWHO + QRNN when correlated against different swarm sizes. The reviewed models GA + QRNN, Adam + QRNN, HOA + QRNN, WHO + QRNN, and the proposed technique AWHO + QRNN when correlated against a swamp size of 5 generated accuracy values of 60.214%, 61.947%, 64.165%, 67.154%, and 71.238%. This illustrates the performance improvement of AWHO + QRNN compared to existing models is by 15.47%, 13.04%, 9.93%, and 5.73%, correspondingly.

#### 4.5 Comparative Techniques

The effectualness of the modulated QRNN\_AWHO is likened to several approaches, like GLAR [7], MLPNN–FFA [5], MFO-Attention-LSTM [12], and ensemble model [14] with varying learning sets and is exemplified below.

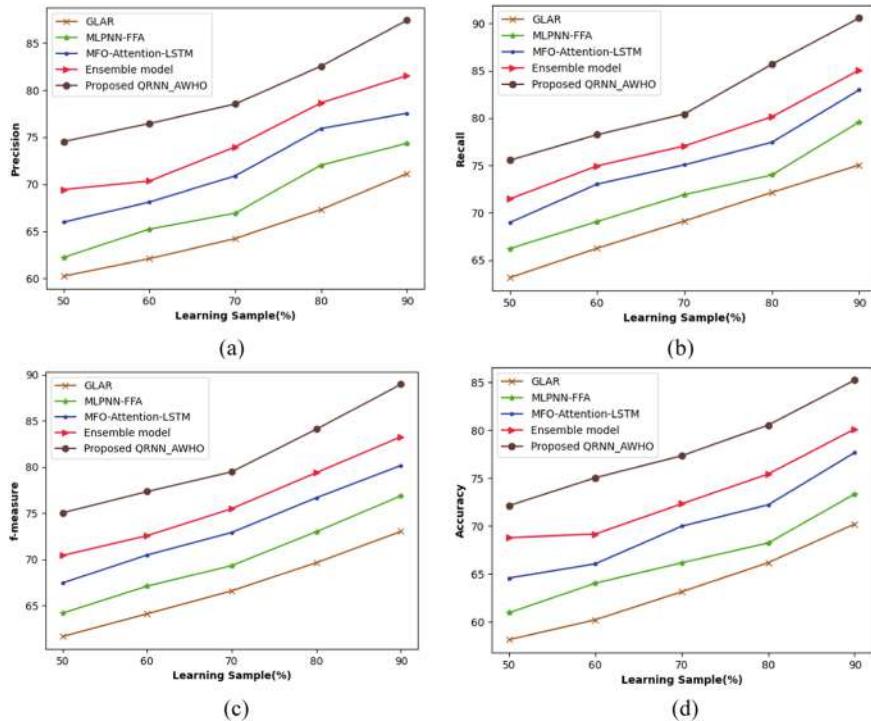
#### 4.6 Comparative Analysis

Figure 4 anticipates the comparative analysis of QRNN\_AWHO relating to the learning samples. Figure 4a portrays the examination of QRNN\_AWHO with respect to precision. When the learning sample is 50%, the reviewed approaches comprising GLAR, MLPNN–FFA, MFO-Attention-LSTM, ensemble model, and the formulated QRNN\_AWHO acquired the precision of 60.234%, 62.237%, 65.981%, 69.432%, and 74.543%. This illustrates the precision improvement of QRNN\_AWHO over existing models as 19.20%, 16.51%, 11.49%, and 6.86%, respectively. Figure 4b exhibits the comparative analysis of QRNN\_AWHO using recall. The recall of the traditional technique GLAR, MLPNN–FFA, MFO-Attention-LSTM, ensemble model, and the established QRNN with AWHO for learning samples 70% is 69.223%,



**Fig. 3** Algorithmic analysis of AWHO + QRNN correlated against varying swarm sizes for **a** precision, **b** recall, **c** F-measure, and **d** accuracy

72.085%, 74.946%, 76.921%, and 80.463%. This demonstrates the recall improvement of QRNN\_AWHO over existing models at 13.97%, 10.41%, 6.86%, and 4.40%, correspondingly. Figure 4c exhibits the comparative evaluation of QRNN-AWHO regarding the F-measure. The F-measure figured when correlated with a learning sample of 90% for the reviewed techniques, like GLAR, MLPNN-FFA, MFO-Attention-LSTM, ensemble model, and proposed technique is 73.174%, 77.057%, 80.259%, 83.188%, and 88.910%. Therefore, the comparative analysis concludes that the proposed QRNN\_AWHO increased its superior performance for F-measure by 17.70%, 13.33%, 9.73%, and 6.44%. Figure 4d exhibits the comparative analysis of the proposed model with accuracy. When correlated against the learning samples for the value 90% and the value attained by the existing models GLAR, MLPNN-FFA, MFO-Attention-LSTM, ensemble model, and proposed technique is 70.212%, 73.345%, 77.675%, 80.126%, 85.234%. The proposed model demonstrated superior accuracy improvements of 17.62%, 13.95%, 8.87%, and 5.99% correspondingly.



**Fig. 4** Comparative analysis of QRNN\_AWHO correlated against different learning samples for **a** precision, **b** recall, **c** F-measure, **d** accuracy

#### 4.7 Comparative Discussion

Table 2 labels the comparative valuation of the proposed QRNN\_AWHO. The proposed QRNN\_AWHO for 90% of learning samples attained maximal precision, recall, and F-measure in the rate 87.433%, 90.565%, and 88.971%. Furthermore, the precision attained by the classical approaches, such as GLAR, MLPNN-FFA, MFO-Attention-LSTM, and ensemble models is 71.130%, 74.354%, 77.545%, and 81.550%. Likewise, the recall achieved by the existing models, like GLAR, MLPNN-FFA, MFO-Attention-LSTM, and ensemble model is 75.043%, 79.588%, 82.965%, and 85.057%. Correspondingly, the F-measure scored by the traditional methods is 73.034%, 76.882%, 80.163%, and 83.266%. This articulated that the AWHO designed to train QRNN to predict the academic performance of students in a blended learning environment exhibited superior performance. AWHO-trained QRNN demonstrates better performance due to its ability to capture long-range dependencies more effectively.

**Table 2** Comparative discussion

Metrics	GLAR	MLPNN–FFA	MFO-attention-LSTM	Ensemble model	Proposed QRNN_AWHO
Precision (%)	71.130	74.354	77.545	81.550	87.433
Recall (%)	75.043	79.588	82.965	85.057	90.565
F-measure (%)	73.034	76.882	80.163	83.266	88.971
Accuracy (%)	70.213	73.345	77.675	80.127	85.234

## 5 Conclusion

The rise in accessibility of Internet Accordance and Collaborative Web Tools and Applications contributed to the development in the huge mass downloads of web services for schools especially in the arena of instigating blended learning. Though executing blending learning is a composite task, the proposed AWHO-trained QRNN aims to seamlessly predict students' performance early through online assessments, facilitating further enhancement of their learning skills with minimal computational resources and time. The academic data from the blended learning medium serves as the initial input. Next, the data normalization is conducted by employing Z-score normalization. Further, the feature selection step is performed through mutual information, and finally, the academic performance prediction is done via employing QRNN trained using the proposed AWHO that is devised by unifying Adam optimizer and WHO. The proposed QRNN\_AWHO is found to be a superior model in students' academic performance prediction that is endowed with the forthcoming highest performance score for the following metrics precision, recall, F-measure, and accuracy at 87.433%, 90.565%, 88.971%, and 85.234%. In further enhancement, hybrid deep learning networks will be integrated to improve the prediction of students' academic performance.

## References

1. MiarNaeimi F, Azizyan G, Rashki M (2021) Horse herd optimization algorithm: a nature-inspired algorithm for high-dimensional optimization problems. *Knowl-Based Syst* 213:106711. <https://doi.org/10.1016/j.knosys.2020.106711>
2. Jais IKM, Ismail AR, Nisa SQ (2019) Adam optimization algorithm for wide and deep neural network. *Knowl Eng Data Sci* 2(1):41–46. <https://doi.org/10.17977/um018v2i12019p41-46>
3. Ali MH, Kamel S, Hassan MH, Tostado-Véliz M, Zawbaa HM (2022) An improved wild horse optimization algorithm for reliability based optimal DG planning of radial distribution networks. *Energy Rep* 8(582):604. <https://doi.org/10.1016/j.egyr.2021.12.023>

4. Johan R, J (2014) Education nowadays. *Int J Educ Sci Res (IJESR)* 4(5):51–56. [https://www.researchgate.net/publication/274704027\\_EDUCATION\\_NOWADAYS#full-text](https://www.researchgate.net/publication/274704027_EDUCATION_NOWADAYS#full-text)
5. Hamadneh NN, Atawneh S, Khan WA, Almejalli KA, Alhomoud A (2022) Using artificial intelligence to predict students' academic performance in blended learning. *Sustainability* 14(18):11642. <https://doi.org/10.3390/su141811642>
6. Al-Faiz MZ, Ibrahim AA, Hadi SM (2018) The effect of Z-Score standardization (normalization) on binary input due the speed of learning in back-propagation neural network" *Iraqi J Inf Commun Technol* 1(3):42–48. <https://www.iasj.net/iasj/download/0a130c438649bd23>
7. Kanetaki Z (2022) Grade prediction modeling in hybrid learning environments for sustainable engineering education. *Sustainability* 14:5205. <https://doi.org/10.3390/su14095205>
8. Chen LQ, Wu MT, Pan LF, Zheng RB (2021) Grade prediction in blended learning using multisource data. *Sci Program* 2021(1):4513610. <https://doi.org/10.1155/2021/4513610>
9. Su F, Zou D, Wang L, Kohnke L (2023) Student engagement and teaching presence in blended learning and emergency remote teaching. *J Comput Educ* 1–26. <https://doi.org/10.1007/s40692-023-00263-1>
10. Müller C, Mildnerger T, Steingruber D (2023) Learning effectiveness of a flexible learning study programme in a blended learning design: why are some courses more effective than others? *Int J Educ Technol High Educ* 20(1):10. <https://doi.org/10.1186/s41239-022-00379-x>
11. Hellas A, Ihantola P, Petersen A, Ajanovski VV, Gutica M, Hyyninen T, Knutas A, Leinonen J, Messom C, Liao SN (2018) Predicting academic performance: a systematic literature review. In: Proceedings of the 23rd annual ACM conference on innovation and technology in computer science education, pp 175–199, July 2018. <https://doi.org/10.1145/3293881.3295783>
12. Qin X, Wang C, Yuan Y, Qi R (2024) Prediction of in-class performance based on MFO-attention-LSTM. *Int J Comput Intell Syst* 17:13. <https://doi.org/10.1007/s44196-023-00395-3>
13. Lu J, Zhao T, Zhang Y (2008) Feature selection based-on genetic algorithm for image annotation. *Knowl-Based Syst* 21(8):887–891. <https://doi.org/10.1016/j.knosys.2008.03.051>
14. Chang W, Cerezo R, Romero C (2021) Multi-source and multimodal data fusion for predicting academic performance in blended learning university courses. *Comput Electr Eng* 89:106908. <https://doi.org/10.1016/j.compeleceng.2020.106908>
15. Kraskov A, Stögbauer H, Grassberger P (2004) Estimating mutual information. *Phys Rev E Stat Nonlinear Soft Matter Phys* 69(6):066138. <https://doi.org/10.1103/PhysRevE.69.066138>
16. Bradbury J, Merity S, Xiong C, Socher R (2016) Quasi-recurrent neural networks. arXiv preprint [arXiv:1611.01576](https://arxiv.org/abs/1611.01576). <https://doi.org/10.48550/arXiv.1611.01576>
17. Academic data derived from blended learning dataset is taken from <https://data.mendeley.com/datasets/z62gdt498/1>. Accessed in July 2024

# A Novel Attention Method to Process Long Trajectories' Sequences Efficiently



Mohammed Abdalla, Hoda M. O. Mokhtar, Abdeltawab Hendawi,  
Tiansheng Yang, and Rajkumar Singh Rathore

**Abstract** Processing users' trajectories has become a crucial task in various aspects of location-based services, such as traffic prediction, trajectory recommendations, tourism recommendations, and travel planning. However, predominately most of the models that currently exist fail when they are trained with long trajectory sequences. This paper proposes a deep learning attention-based model, named (*SAMO*) to efficiently process the long trajectory sequences of moving objects efficiently on road networks. Indeed, (*SAMO*) stands for Spatial Attention Model for Objects' Movements. The proposed model promises to process the very long trajectory sequences that will most likely be visited by the moving object whether the object's self-history is available. In the case of no self-history, the model starts by catching the  $k$  nearest moving objects in the vicinity. In the case of self-history, the model selects trajectories like its current trip from stored history. After that, the model is trained by these objects' trajectories either  $k$  nearest objects or similar objects and then focuses on the significant parts of these trajectories to generate results of processing and analysis of the trajectory. Overall, the proposed model outperforms competitive models by achieving up to 98% accuracy for the next multi-step prediction.

---

M. Abdalla (✉)

Faculty of Computers and Artificial Intelligence, Beni-Suef University, Beni Suef, Egypt  
e-mail: [mohammed.a.youssif@fcis.bsu.edu.eg](mailto:mohammed.a.youssif@fcis.bsu.edu.eg)

H. M. O. Mokhtar

Egypt University of Informatics, Cairo, Egypt  
e-mail: [hoda.mokhtar@eui.edu.eg](mailto:hoda.mokhtar@eui.edu.eg)

A. Hendawi

Department of Computer Science and Statistics, University of Rhode Island, Kingston, USA  
e-mail: [hendawi@uri.edu](mailto:hendawi@uri.edu)

T. Yang

University of South Wales, Llantwit Rd, Pontypridd, United Kingdom  
e-mail: [tiansheng.yang1@southwales.ac.uk](mailto:tiansheng.yang1@southwales.ac.uk)

R. S. Rathore

Cardiff School of Technologies, Cardiff Metropolitan University,  
Llandaff Campus, Western Avenue, Cardiff, United Kingdom  
e-mail: [rsrathore@cardiffmet.ac.uk](mailto:rsrathore@cardiffmet.ac.uk)

**Keywords** Deep learning · Moving objects · Neural network · Trajectory analysis · Attention model · Transportation sustainability

## 1 Introduction

Location-based services (LBSs) play a critical role in many businesses and organizations. These services mainly depend on location-related data. The extracted patterns from this location-related data can give us powerful insights to better understand strong spatial relationships [1]. According to [2], over 2.5 billion users are predicted to use smartphones worldwide by the end of 2019. This means that more people will be moving with location-enabled devices in their pockets, bolstering the significance and potential of LBS.

Processing the trajectories of an end-user helps LBS to anticipate the user's needs, which aids planning. LBS mainly depends on these trajectories to complete the analytic process that is used in different use cases like planning trips, predicting next routes, navigation services, and so on. By processing this data using machine learning techniques, it is relatively easy to process the trajectories of moving objects and make the best benefits from them.

To better understand human mobility, several models have been developed to process the trajectories of moving objects [3–6]. However, these models necessitate the existence of the self-historical motions of these objects to process trajectories efficiently. Additionally, these models perform poorly when trained on long sequences to generate predictions. Currently, these models suffer from more issues such as: (1) not being able to anticipate the upcoming motions when the self-history of the query object is insufficient, (2) depending on false assumptions like linear motion or following the shortest trajectory, and (3) performing poorly when the model is trained with very long trajectories.

To remedy these issues, this paper proposes an attention model that processes efficiently the trajectories of the moving object. The overarching goal of this paper can be accomplished through the pursuit of two specific research objectives: (1) accurately process long trajectories in a way that provides a better understanding of human mobility, and (2) perform well when the learning model is trained by very long trajectories.

Indeed, *SAMO* discriminates from other conventional trajectory models by obviating the way these models are used to process trajectories by scanning the whole input trajectory sequence to generate insights into human mobility. To generate the outputs, the *SAMO* only considers the important points of the input trajectory sequences. As a result, *SAMO* enables the internal representation of input trajectories to be improved in light of pertinent data obtained from the query object. Then, by compiling data relevant to the final representation, only the information required to forecast the query object's eventual answer is given.

**Contribution.** The following are this paper's main contributions:

- We provide a deep learning model called *SAMO* for processing the trajectories of a moving object. When the query object's self-history is insufficient, *SAMO* is regarded as an innovative attempt that uses attention models to process extended trajectories.
- We propose a projection layer that dynamically chooses the input features; this layer confirms that our model works efficiently when the dataset size increases. This layer also significantly improves the *SAMO* efficiency by preventing overfitting [7] problems during the prediction phase and by disregarding anomalies in the input data.
- We ensure model performance optimization when dealing with long input sequences.

**Roadmap.** This is how the remainder of the paper is structured. A brief overview of current encoder–decoder architectures and their interactions with attention models is provided in Sect. 2. Trajectory prediction-related research and alternative approaches are discussed in Sect. 3. Section 4 describes the suggested strategy. Our model is experimentally evaluated in Sect. 5. The paper is finally concluded in Sect. 6.

## 2 Related Work

Several attempts were made to address the issue of trajectory processing. Each of these distinctively tackled the problem. However, in their solutions, they mainly learned by using one of three models; machine learning or deep learning or machine learning or deep learning with attention. This section reviews those attempts.

### 2.1 Machine Learning Models in Trajectory Processing

There are several methods designed to process trajectories of moving objects according to the Hidden Markov Model (HMM) and frequent pattern matching [4, 8–14]. In [10], the authors develop a similarity-based prediction algorithm to deduce the end-to-end trajectory of a vehicle based on the vehicle's past trips' observations. In [13], a simple Markov model is used to predict the short-term trajectory for the drivers. Authors in [11, 14] employed a variable-order Markov model to predict future trajectories and consider dynamic traffic conditions. In [4, 12], processing techniques are developed to detect frequent trajectory latent patterns, employing these patterns to predict the most possible location of moving objects. In [15], authors propose an algorithm to execute a predictive query in the road network and return the top-k available moving objects, and this algorithm is empowered by an index to efficiently access queries in the road network with a large number of moving objects.

## 2.2 Deep Learning Models in Trajectory Processing

There are several techniques and approaches proposed to process the trajectories of moving objects according to neural network concepts [16–23]. These approaches deploy recurrent cell-type Long Short-Term Memory (LSTM) to process trajectories [16, 17, 20, 23]. Authors in [18] propose a novel method named *DeepMove* that models the movements between places and generates statistical results. In [20], the authors introduce a novel approach called T-CONV which captures and then models the trajectories of moving objects as two-dimensional images and then extracts areas with a high impact on the final prediction. In [16], the authors propose a neural network model that predicts the destination of a taxi based on trajectory prefix and associated meta-data, using a recurrent bidirectional neural network to encode the prefix and meta-data embeddings.

## 2.3 Deep Learning Attention Models in Trajectory Processing

Authors in [24] propose a novel method to process the motion of a pedestrian given a short history of their, and their neighbors' past behavior. The proposed method is the combined attention model which utilizes both soft attention to map the trajectory information from the history to the future positions of the pedestrian of interest. In [25], the authors proposed using attention mechanisms to incorporate network traffic state data into urban vehicle trajectory prediction. Authors in [26] propose an end-to-end deep learning model to learn the movement patterns of humans using different navigational modes directly from data using the much popular sequence-to-sequence model coupled with a soft attention mechanism.

*SAMO* differentiates itself from the above studies by being the attempt that perform the trajectory processing without considering the whole trajectory sequence in the training or learning cycles and focusing only on the vital parts of the source input trajectory to produce the results. Additionally, the proposed model considered building models also based on similar datasets which make processing work in an efficient manner and prevent learning models from failing under overfitting conditions.

## 3 Background

This section provides the necessary background information that will be used throughout the rest of the paper.

**Table 1** Paper acronyms

RNN	Recurrent neural network
$Q_\tau$	Query object's trajectory
KNN	K-nearest neighbors
LSTM	RNN type stands for Long Short-Term Memory
GRU	RNN type stands for gated recurrent units

### 3.1 Attention Model Overview

The encoder and decoder recurrent neural network architecture [27, 28] is an architecture that encodes the input sequences into a fixed-length vector, then reads this vector, and decodes it into an output sequence. Although the encoder–decoder architecture achieves excellent results on various problems, it suffers from the strong limitation that all input sequences are forced to be encoded to a fixed-length vector. The drawbacks of this architecture are limiting the performance of neural networks, particularly when the model has long input sequences to train because a neural network needs to scan all the information of a source sequence, and this is difficult for the neural network to cope with long sequences, especially those that are longer than the sequences in the training corpus [29]. Paying attention within sequences [30, 31] is the idea of removing the fixed-length representation constraint from the encoder–decoder architecture. This is done by saving the intermediate results from the encoder at each step of the input sequence, training the model to pay attention to these inputs, and binding them to items in the output sequence. As a result, each item in the output sequence depends on selective items in the input sequence. Each time the model produces a segment in output; it searches for a set of positions in a source input sequence where the most relevant information is concentrated. Then, the model predicts a target segment based on context vectors associated with these positions and all the previously produced target segments.

Table 1 summarizes all acronyms that will be used throughout the paper.

### 3.2 Attention Model Building

The perdition function deployed inside *SAMO* is built based on attention model notions. Thus, *SAMO* starts the training on multiple trajectories at once determined by the (*batch\_size*) parameter. For each iteration, the *SAMO* will be fed by the following parameters: (1) the *encoder\_input*, which represents the source input tokens (tokens before tab space); (2) the *decoder\_input*, which represents the target input tokens (tokens after tab space); and (3) the *decoder\_output*, which represents the target output tokens; the *decoder\_inputs* are shifted to the left by one-time step *t* with an end-of-sequence tag appended on the right.

Before the training of the model, the following parameters must be configured:

- *rnn\_unit*, this parameter characterizes the count of units per layer on encoder and decoder.
- *rnn\_cell*, this parameter characterizes the RNN cell type of encoder and decoder. The RNN cell type includes LSTM or GRU. The main difference between LSTM [32] and GRU [33] is that LSTM's architecture has three gates and GRU has only two gates; additionally, the update gate in GRU fulfills the roles of input and forget gate in LSTM. As a result, GRU might take less time in training than LSTM.
- *encoder\_rnn\_type*, this parameter characterizes the encoder's RNN type. The RNN type can be unidirectional or bidirectional [34–36]. The major difference between unidirectional and bidirectional representation is that the unidirectional representation learns the sequences from the previous time-stamps only; however, the bidirectional representation learns the sequences from both previous and future time-stamps. As a result, bidirectional better understands the context more than unidirectional representations.
- *attention\_mechanism*, this parameter characterizes the attention mechanism used during the training of the model.
- *dropout\_keep\_prob\_rnn*, this float parameter characterizes the dropout keep probability on RNN ( $> 0.0, \leq 1.0$ ). Commonly, dropout [37–39] is a regularization method where input, output, and recurrent connections to RNN units are ignored by certain probability from activation and weight updates during the network training phases; this has a great impact on reducing the occurrence percentage of overfitting problem and improving the model performance.

The proposed model computes the attention at every decoder time step as illustrated in Fig. 1. The attention computation is accomplished by the following steps:

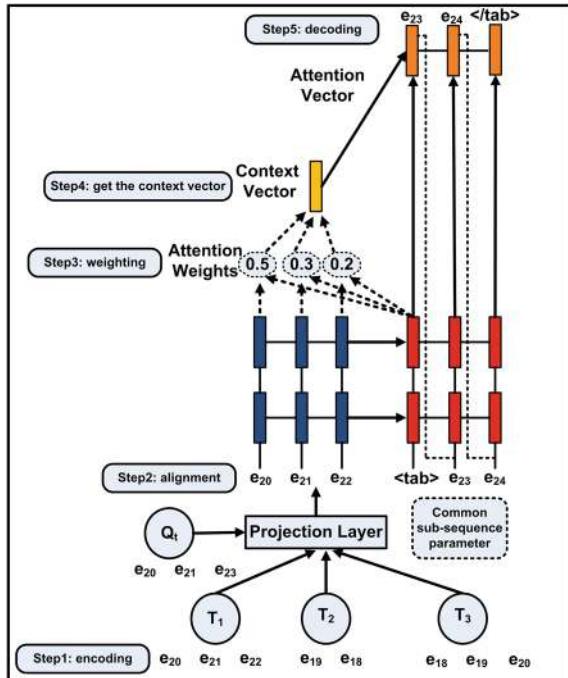
- Compare the current target hidden state with all source states to obtain the attention weights.
- Then, based on the attention weights, the context vector is computed as the weighted average of the source states.
- Next, the context vector is combined with the current target hidden state to obtain the final attention vector.
- Finally, the attention vector is fed as an input to the next time step. These steps can be summarized by the following key equations:

$$\alpha_{ts} = \frac{\exp(\text{score}(h_t, \bar{h}_s))}{\sum_{k=1}^s \exp(\text{score}(h_t, \bar{h}_k))} \quad [\text{Attention weights}] \quad (1)$$

$$c_t = \Sigma \alpha_{tk} \bar{h}_k \quad [\text{Context vector}] \quad (2)$$

$$a_t = f(c_t, h_t) = \tanh(W_c[c_t, h_t]) \quad [\text{Attention vector}] \quad (3)$$

**Fig. 1** Spatial attention mechanism example



### 3.3 Summary

To sum up, an input sequence is encoded into a sequence of vectors and then chooses a subset of these vectors adaptively when decoding the output. As a result, this will make a neural network free from squashing all the information of a source sequence regardless of the input sequence length.

## 4 Problem Definition

Let  $Q_\tau$  represent the trajectory of a query moving object; this can be expressed as an ordered list of edges that the query object  $Q_\tau = (e_1, \dots, e_n)$  has visited. Let  $S$  be the set of trajectories of other moving objects that are presently moving in tandem with the query object;  $S = \{\tau_1, \dots, \tau_{|S|}\}$ ;  $\tau_i$  is defined as an edge sequence. The number of next moves that the query object seeks to process is  $\mathcal{F}$ . When the motion segments of the query object are quite long, our goal is to analyze  $\mathcal{F}$  motion segments given  $Q_\tau, S$ , and  $\mathcal{F}$ . The primary goal is to accomplish effective processing so that the model may be trained with lengthy sequences without interruption.

## 5 Proposed Model

This section illustrates the proposed model (*SAMO*).

### 5.1 Main Idea

The main idea beyond *SAMO* is to handle long sequence trajectories of a moving object without performance downgrade in the learning model. First, *SAMO* searches for the moving object's historical trips, if *SAMO* does not find the historical trips, then *SAMO* identifies the current location point of the moving object (longitude, latitude) and then tracks the  $k$  nearest moving objects within a specific distance  $d$  using the range search query. As a result, *SAMO* extracts the  $K$  nearest moving objects trajectories and considers them as the input features that will be used during the training process. On the other hand, if *SAMO* finds the self-history of the moving object, it gets similar trajectories that are the same as the moving object in its current trip. As a result, *SAMO* selects these similar trajectories and considers them as the input features that will be used during the training process. *SAMO* adopted its processing module based on attention-learning notions. So, before starting the training phase, *SAMO* converts each input trajectory to a string data type and splits each trajectory into multiple tokens; each token represents the road segment. The primary pre-requisite for encoding/decoding operations that run inside the attention model is dividing the input trajectory string into two parts separated by tab spaces; tab spaces make the model switch the transition mode from encoding to decoding. After that, the model is trained based on these modified strings and returns as an output a hash-map data structure that contains the trajectory sequences as keys and possible target trajectory sequences with probabilities as values. Finally, the *SAMO* returned a hash-map (key/value) pair of constructed segments and other segments they are following with probability percentage as an output of processing. The constructed hash-map result can be used later in activities like planning roads, route predictions, and routing services.

### 5.2 SAMO Performance Tuning: Projection Layer

To make our model scalable and able to accept the higher load, we developed a tuning parameter called a *common sub-sequence match*. This parameter works to filter out the input sequences that does not match the configured threshold value. The major gains of this tuning are the following: (1) avoid the model distraction, (2) save computation cost, and (3) ensure that the model will work effectively when datasets become very large. The *common sub-sequence match* tuning parameter exists into the employed projection layer of our model as shown in Fig. 1. For example assume

the query object's trajectory  $Q_\tau$  ( $e_{20}, e_{21}, e_{23}$ ) and other trajectory sequences are  $\tau_1(e_{20}, e_{21}, e_{22})$ ,  $\tau_2(e_{19}, e_{18})$ , and  $\tau_3(e_{18}, e_{19}, e_{20})$ . Assume that we configured the common sub-sequence match threshold as 0.3. As a result, the model will filter out  $\tau_2$  and accept  $\tau_1$  and  $\tau_3$  as input sequences for further processing. The common sub-sequence match is computed based on the following equation:

$$\text{SequenceMatch} = \frac{\text{Length}(\text{common sequence}(Q_\tau, \tau_i))}{\text{Length}(Q_\tau)} \quad (4)$$

### 5.3 R-Tree: Index for Spatial Search

In the case of no self-history of the query object, *SAMO* exploits R-Tree to get  $k$  nearest objects surrounding the query object within a specific distance. R-Tree [40] is considered as one of the most powerful access techniques in the area of spatial data management. Indeed, R-Tree is designed for answering a range of queries and returns moving objects inside this range. Moreover, R-Tree represents each object by a minimum bounding d-dimensional rectangle (MBR). The root of the tree represents an MBR that includes all objects indexed by the tree and each node corresponds to the MBR that bounds its children. Consequently, R-Tree answers the range queries by traversing the tree starting from the root and ending at the leaves, accessing only nodes whose MBRs intersect with the query range.

### 5.4 Algorithm

The pseudo-code of *SAMO* is demonstrated in Algorithm 1. The algorithm requires five input parameters: (a) the trajectory of the query item; (b) the trajectories of currently moving objects; (c) the number of previously moved edges,  $\lambda$ ; (d) distance  $d$ , which represents the specific distance which  $Q_\tau$  needs to get nearest moving objects within this distance; and (e)  $\text{SequenceMatch}_{\text{Threshold}}$  which represents the number of tokens shared between  $Q_\tau$  and other input trajectories' sequences. The output is the most probable next segments that  $Q_\tau$  will follow. In line 9, the algorithm checks if the query object has no self-history, then the algorithm gets the longitude and latitude of  $Q_\tau$  current edge (line 11). Then, in line 13 the algorithm uses R-Tree to get the  $K$  nearest moving objects within distance  $d$  according to  $Q_\tau$  current location from objects currently moving in the system  $Trajectories_{Current}$ , after that the algorithm stores these objects as strings in  $KNN_{List}$ . It is critical to note that  $KNN_{List}$  is always  $\subseteq Trajectories_{Current}$ . After that, the algorithm iterates over each trajectory string included in  $KNN_{List}$ , for each iteration the algorithm computes the match score between query object trajectory  $Q_\tau$  and the iterated one,

**Algorithm 1** Spatial Attention Model for Trajectory Prediction

---

```

1: INPUT: Query object's trajectory  $Q_\tau$ , Current moving objects trajectories  $Trajectories_{Current}$ ,
   Number of Previous Moved Edges  $\lambda$ , Distance  $d$ ,  $SequenceMatchThreshold$ 
2: SET  $QueryObjectLocationPoint Q_{Point(x,y)} \leftarrow \phi$ 
3: /* The current edge of the query object */
4: SET  $QueryObjectLocationPoint Q_{CurrentEdge} \leftarrow \phi$ 
5: /* This list stores nearest objects of  $Q_\tau$  within  $d$  */
6: SET  $NearestneighborList KNN_{List} \leftarrow \phi$ 
7: /* This list stores similar trajectories like  $Q_\tau$  */
8: SET  $SimilarTrajectories S \leftarrow \phi$ 
9: if  $Q_\tau$  has no self-history then
10:   /* Get the longitude and latitude of  $Q_\tau$  current edge */
11:    $Q_{Point(x,y)} \leftarrow$  Get the location of  $Q_{CurrentEdge}$ 
12:   /* Store trajectories of K objects as strings in  $KNN_{List}$  */
13:    $KNN_{List} =$  get  $K$  nearest objects from  $Q_{Point(x,y)}$  within  $d$ 
14:   for each string  $\tau_i$  in  $KNN_{List}$  do
15:      $MatchScore = SequenceMatch(Q_\tau, \tau_i)$ 
16:     if  $MatchScore \leq SequenceMatchThreshold$  then
17:       Remove  $\tau_i$  from  $KNN_{List}$ 
18:     end if
19:   end for
20:   Build the Attention model from  $KNN_{List}$ 
21: end if
22: if  $Q_\tau$  has self-history then
23:   for each  $\tau_i$  in  $Q_\tau$  Previous History do
24:      $MatchScore = SequenceMatch(Q_\tau, \tau_i)$ 
25:     if  $MatchScore = 1$  then
26:       add  $\tau_i$  in  $S$ 
27:     end if
28:   end for
29:   Build the Attention model from  $S$ 
30: end if
31:  $QueryResult =$  Query attention model by  $\lambda$  tokens of  $Q_\tau$ 
32: OUTPUT: Return  $QueryResult$ 

```

---

and if it is less than  $SequenceMatchThreshold$  defined by the projection layer, the algorithm will remove it from  $KNN_{List}$ , otherwise the algorithm keeps it (lines 9–21). At the end, the algorithm trains the attention model based on created  $KNN_{List}$ . In line 22, the algorithm checks if the query object  $Q_\tau$  has self-history. Then, the algorithm iterates over the saved previous history of the  $Q_\tau$ . After that, the algorithm gets from the history only trajectories that moved exactly like the  $Q_\tau$ , and this is done by checking that the sequence match value  $MatchScore$  is equal to 1. If the  $MatchScore$  between  $Q_\tau$  and iterated trajectory is 1, then it is saved in similar trajectories list  $S$ , otherwise, it is not saved. At the end, the algorithm trains the attention model based on created  $S$  list, (lines 23–29). The output at the end of the training phase is a hash-map (key/value) pairs. The algorithm gets  $\lambda$  tokens moved by the  $Q_\tau$ , then *SAMO* query the created hash-map and returns the next segments that the  $Q_\tau$  can be follow in future.

## 6 Experiments

This section evaluates experimentally the proposed *SAMO* model.

### 6.1 *Experiments Objectives*

The objectives of our experiments are to prove that *SAMO* is robust against baseline models, in addition, to reach for the ideal values that must be configured for the model to make it performs well and generates accurate predictions.

### 6.2 *Experimental Setup*

#### 6.2.1 Datasets

The tests carried out in this study make use of real-life trajectories that were collected between April 2007 and August 2012 by Microsoft Research as part of the Geo-life project [41]. Additionally, these trajectories are separated into smaller trajectories, each consisting of 10 average-length road segments. Consequently, there are 2500 paths in all. Every trajectory is split into two parts based on the specified value of  $\lambda$ , which in our experiments is 4, so, each trajectory is split into two parts separated by a tab space: the first consists of 4 segments (tokens) and the second part consists of the rest of the segments (6 segments). All trajectories GPS points (longitude, latitude) are map-matched to road segments (edges) over the road network. The road network data is captured from OpenStreetMap [42]. Road network data in this work demonstrates Hamilton city in the USA. The map-matching algorithm is out the scope of this paper [43].

#### 6.2.2 Experimental Settings

Tensorflow 1.4.0 and Python are used to implement the prediction function used in *SAMO*. A well-known open-source Python library designed for applications requiring extensive computations is called Tensorflow [44]. The training parameters indicated in Table 2 are used to train *SAMO*. Java with JDK 1.9 is used to implement the R-Tree module within Eclipse PHOTONID. Every experiment is carried out on a Windows 10 PC equipped with an Intel(R) Core(TM) i7 CPU and 24GB of RAM.

**Table 2** Model hyperparameters

Parameter	Value
Train epochs	5
Number of layers	3
Train steps	2500
Train batch size	512
RNN unit	1024
RNN cell	LSTM
Encoder rnn type	Bidirectional
Dropout keep prob rnn	0.8
Attention mechanism	Bahdanau
Optimizer	Adam
Learning rate	0.001

### 6.2.3 Criteria for Evaluation

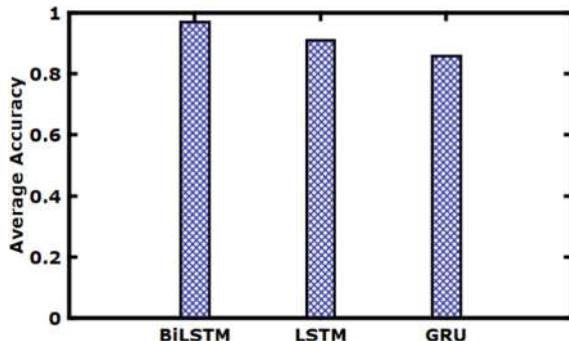
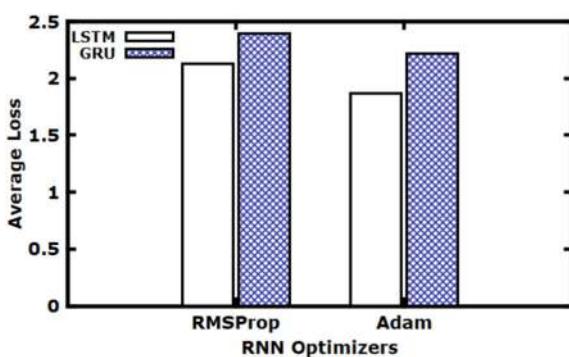
Three criteria were used in this study to assess our model *SAMO* at different stages: average loss, average accuracy, and CPU processing time. The total sum of mistakes connected to every sample in the training dataset is indicated by the average loss value. The model's ability to make accurate predictions is indicated by the average accuracy value. The average amount of time needed to generate the final result is implied by the CPU processing time.

## 6.3 Adjustment Model Hyperparameters

This section explores the adjustment of hyperparameters that *SAMO* needs to be trained on them to obtain the best accuracy achievement. Table 2 illustrates these parameters. Next, we present our research results.

**Exp1:-Impact of choosing RNN cell type.** The average accuracy of three distinct RNN cell types—Bidirectional LSTM (BiLSTM) units, LSTM, and GRU—is compared in this series of tests by Fig. 2. The RNN types are displayed on the X-axis in Fig. 2. From 0 to 1, the accuracy numbers are displayed on the Y-axis. Compared to other varieties, BiLSTM is reported to produce higher accuracy. BiLSTM's capacity to manage training over datasets in both directions provides a strong understanding of the context and eliminates ambiguity, which serves as an explanation for this. Furthermore, the LSTM cell type is observed to acquire greater accuracy compared to GRU. In summary, accuracy is increased by 8% greater by BiLSTM than by LSTM.

**Exp2:-Impact of choosing optimizer.** In this set of experiments, Fig. 3 demonstrates the average loss value resulted from optimizers; RMSProp and Adam. This set of

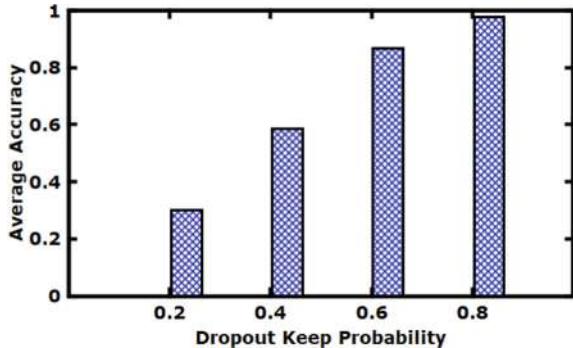
**Fig. 2** RNN cell types**Fig. 3** Optimizers

experiments tests the choosing of optimizer against RNN cell types; LSTM and GRU. The Adam optimizer stands for Adaptive Moment Estimation, and it uses adaptive learning rates to converge faster. Adam starts off with big steps and finishes with small steps, it moves faster initially as the learning rate decay, smaller and smaller steps, allowing the model to converge faster since it doesn't overstep the local minimum with as big steps. While RMSProp optimizer stands for Root Mean Squared Propagation, the important property of RMSProp is that it is not restricted to just the sum of the past gradients, but instead, it is more restricted to gradients for the recent time steps.

The RNN optimizers are displayed in Fig. 3, with the X-axis representing them. The loss values range from 0 to 2.5 on the Y-axis. Adam clearly outperforms RMSProp in terms of loss values, as seen by the latter's lower achievement. In conclusion, compared to RMSProp optimizer, Adam optimizer reduces the loss value by about 0.4.

**Exp3:-Impact of increasing dropout keep probability.** The average accuracy of various dropout maintain probabilities in this collection of tests is compared in Fig. 4. Dropout keep probability means that if the keep probability is 0.8, this means that only 0.2 of the network nodes will be ignored during the training and 0.8 of the network is kept. As illustrated in Fig. 4, it is observed that with increasing the dropout keep

**Fig. 4** Dropout keep probability



probability, there is an increase in average accuracy. However, from the practice, the dropout keep probability should ideally be 0.8 to achieve the local minima (best accuracy). In summary, increasing the dropout keep probability by 0.2 will improve the average accuracy by 20%.

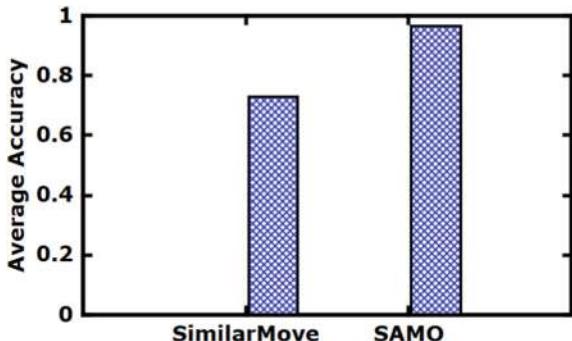
#### 6.4 Comparative Analysis with Baseline Approaches

This section compares the *SAMO* with baseline approaches that predict the next steps of moving objects. We next report our findings.

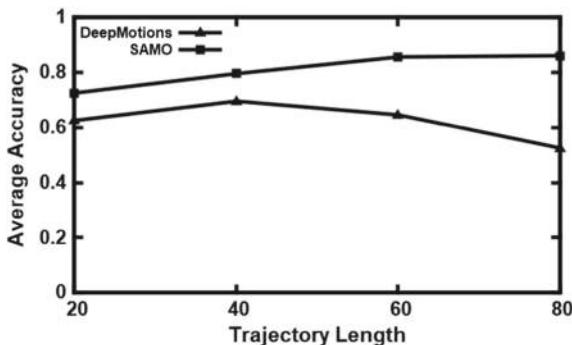
**Exp4:-Impact of choosing prediction function.** In this series of tests, Fig. 5 contrasts *SimilarMove* with *SAMO* to forecast an object's upcoming move, *SimilarMove* [45] is a system that predicts the future movements of moving objects on road networks without relying on their past trajectories, and it depends in its work on Hidden Markov Model (HMM) in the learning process, while *SAMO* depends in its work in deep learning attention-based learning model in the learning process. The Y-axis shows the accuracy as a value between 0 and 1. As shown in Fig. 5, it is observed that the *SAMO* achieves higher accuracy than the *SimilarMove* because the *SAMO* understands well the context and focuses only into the most important parts of input sequences. To conclude, *SAMO* achieves accurate predictions of more than 22% compared to *SimilarMove*.

**Exp5:-Impact of increasing the trajectory length.** Figure 6 compares *SAMO* with *DeepMotions* system to examine the effect of increasing the length of the trajectory on the prediction accuracy, *DeepMotions* [46] is a prediction system that uses deep learning to forecast a moving query object's future path. Figure 6 reveals that in the case of an increasing number of segments in the trajectory and generates the predictions using *DeepMotions* the accuracy decreases at some level, while when using *SAMO* attention-based deep learning model, it is observed that the accuracy does not impact by the trajectory length and additionally. This experiment proves that *SAMO* performs efficiently with long trajectories sequences.

**Fig. 5** Impact of choosing prediction function



**Fig. 6** Impact of increasing the trajectory length



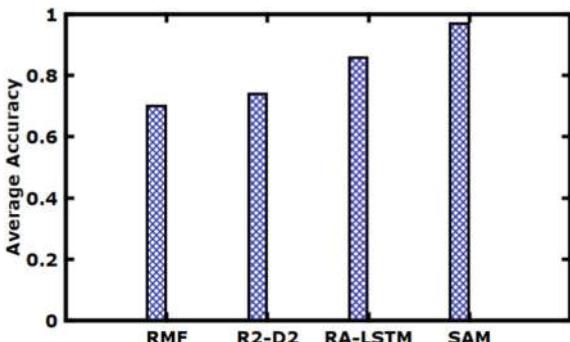
**Exp6:-Competitive methods comparison.** In this set of experiments, Fig. 7 compares 3 baseline approaches of trajectory prediction by *SAMO* as follows:

- RMF [47], this trajectory prediction approach is based on models and computes a motion function to record motions.
- R2-D2 [48], it is a probabilistic model that forecasts future trajectory and is based on HMM.
- RA-LSTM [49], it is a neural network model that solves multi-step vehicle trajectory prediction tasks by combining road-aware features.

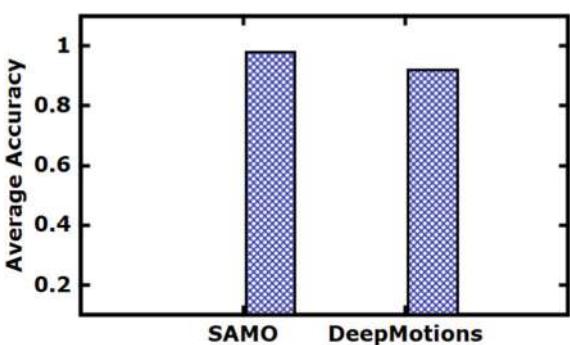
As shown, it is observed that *SAMO* outperforms the competitive prediction methods by a range from 5–20% in average accuracy.

**Exp7:-Impact of paying attention.** In this set of experiments, Fig. 8 compares the average accuracy between *DeepMotions* and *SAMO*. *DeepMotions* is a deep learning model that predicts the next movements of moving objects, while *SAMO* is a deep learning attention-based model that predicts the next movements of moving objects. It is observed that *SAMO* achieves higher accuracy more than *DeepMotions* in predicting the next 5 steps of moving objects, and *SAMO* outperforms *DeepMotions* by 5% in the average accuracy. The justification behind this is that *SAMO* learns where to focus on the significant parts in the input sequences for each token, and this makes the model more robust and ensures the run time optimization of the model.

**Fig. 7** Comparison between competitive methods



**Fig. 8** RNN with and without attention



## 7 Conclusion

In this work, we have designed a spatial attention model for trajectory prediction named *SAMO*. The primary goal of *SAMO* is to correctly forecast a query object's future trajectories in the absence of its self-history. It excels when trained with lengthy sequences. *SAMO* is considered as an attempt that employs attention encoder-decoder architecture for trajectory prediction purposes. *SAMO* is evaluated against different evaluation metrics and in comparison to other baseline models. Experiments proved that *SAMO* achieves significant improvements in prediction accuracy and works efficiently with very long trajectory sequences. We intend to develop a scalable framework in the future work that enables end users to submit predictive queries via big data frameworks on the cloud. We also look at using GPU microprocessors to improve performance and effectively handle several jobs at once.

## References

1. Location-based-services-are-playing-a-crucial-role-in-mobile-app development. <http://www.vensi.com/>, Accessed on April 2019
2. Importance-of-location-based marketing. <http://tickto.com/>. Accessed on April 2019
3. Karimi H, Liu X (2003) A predictive location model for location-based services. In: Proceedings of the ACM-GIS'03
4. Shen H, Jeung H, Liu Q, Zhou X (2008) A hybrid prediction model for moving objects. In: 2008 IEEE 24th international conference on data engineering
5. Song C, Qu Z, Blumm N, Barabási A-L (2010) Limits of predictability in human mobility. Science 327(5968)
6. Sun J, Papadias D, Tao Y, Liu B (2004) Querying about the past, the present, and the future in spatio-temporal databases. In: Proceedings 20th international conference on data engineering
7. Webb GI (2017) Overfitting. Encyclopedia of machine learning and data mining, pp 947–948
8. Abdalla M, Islam A, Ali M, Hendawi A (2024) Framework to process vehicles uncertain locations for intelligent transportation. Int Dec Technol 18:1–17
9. Abdalla M, Mokhtar HMO, Elgamal N (2020) HarmonyMoves: a unified prediction approach for moving object future path. Int J Adv Comput Sci Appl (IJACSA)
10. Froehlich J, Krumm J (2008) Route prediction from trip observations. SAE Technical Paper Series
11. Necula E (2014) Dynamic traffic flow prediction based on GPS Data. In: 2014 IEEE 26th international conference on tools with artificial intelligence
12. Qiao S, Han N, Zhu W, Gutierrez L (2015) TraPlan: an effective three-in-one trajectory-prediction model in transportation networks. IEEE Trans Intell Transp Syst 16(3) 2015
13. Simmons R, Browning B, Zhang Y, Sadekar V (2006) Learning to predict driver route and destination intent. In: 2006 IEEE intelligent transportation systems conference
14. Zheng Y, Wang Q, W, Kuang M (2008) Research into the driver's route choice under existing real-time traffic information. In: 2008 IEEE international conference on Industrial Engineering and Engineering Management
15. Yoon S-H, Park S, Kim T, Kim JW, Park S (2019) Predictive query processing considering the movement of both user and objects. In: 2019 IEEE international conference on big data and smart computing (BigComp)
16. de Alexandre B, Étienne S, Alex A (2015) Pascal Vincent14, and Yoshua Bengio14. Artificial neural networks applied to taxi destination prediction, ECML-PKDD-DCs
17. Fathollahi M, Kasturi R (2016) Autonomous driving challenge: to infer the property of a dynamic object based on its motion pattern. Lecture notes in computer science computer vision—ECCV 2016 Workshops, pp 40–46
18. Feng J, Li Y, Zhang C, Sun F, Meng F, Guo A, Jin D (2018) DeepMove: predicting human mobility with attentional recurrent networks. In: Proceedings of the 2018 world wide web conference on world wide web—WWW 18
19. Kim B, Kang CM, Kim J, Lee SH, Chung CC, Choi JW (2017) Probabilistic vehicle trajectory prediction over occupancy grid map via recurrent neural network. In: IEEE 20th international conference on intelligent transportation systems (ITSC)
20. Lv J, Li Q, Sun Q (2018) T-CONV: a convolutional neural network for multi-scale taxi trajectory prediction. In: IEEE International conference on big data and smart computing (BigComp)
21. Mikolov T, Karafli M, Burget L, Cernocký J, Khudanpur S (2010) Recurrent neural network based language model. In: Proceedings of the conference of the international speech communication association (INTERSPEECH 2010), pp 1045–1048
22. Wang Liu SL, Wu Q, Tan T (2016) Predicting the next location: a recurrent model with spatial and temporal contexts. In: Proceedings of the thirtieth AAAI conference on artificial intelligence, Phoenix, pp 194–200
23. Wu F, Fu K, Wang Y, Xiao Z, Fu X (2017) A spatial-temporal-semantic neural network algorithm for location prediction on moving objects. Algorithms, 10(2)

24. Fernando T, Denman S, Sridharan S, Fookes C (2018) Soft + hardwired attention: an LSTM framework for human trajectory prediction and abnormal event detection. *Neural Netw* 466–478
25. Choi S, Kim J, Yeo H (2019) Attention-based recurrent neural network for urban vehicle trajectory prediction. *Proc Comput Sci* 327–334
26. Varshneya D, Srinivasaraghavan G (2017) Human trajectory prediction using spatially aware deep attention models. In: 31st Conference on neural information processing systems
27. Cho K, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)
28. Sutskever I, Vinyals O, Le QVV (2014) Sequence to sequence learning with neural networks. In NIPS pp 3104–3112
29. Cho K, van Merriënboer B, Bahdanau D, Bengio Y (2014) On the properties of neural machine translation: encoder-decoder approaches. In: In eighth workshop on syntax, semantics and structure in statistical translation
30. Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. ArXiv 2014
31. Luong M-T, Pham H, Manning CD (2015) Effective approaches to attention-based neural machine translation. In: Proceedings of the 2015 conference on empirical methods in natural language processing
32. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 1735–1780
33. Chun J, Gulcehre C, Cho K, Bengio Y (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. ArXiv, 2014
34. Schuster M, Paliwal KK (1997) Bidirectional recurrent neural networks. *IEEE Trans Sig Proc*
35. Singh B, Marks TK, Jones M, Tuzel O, Shao M (2016) A multi-stream Bi-directional recurrent neural network for fine-grained action detection. In: IEEE conference on computer vision and pattern recognition (CVPR)
36. Vu NT, Gupta P, Adel H, Schtze H (2016) Bi-directional recurrent neural network with ranking loss for spoken language understanding. In: IEEE international conference on acoustics, speech and signal processing (ICASSP)
37. Bayer J, Osendorfer C, Korhammer D, Chen N, Urban S, van der Smagt P (2013) On fast dropout and its applicability to recurrent networks. ArXiv preprint [arXiv:1311.0701](https://arxiv.org/abs/1311.0701)
38. Haşim Sak AS, Beaufays F (2014) Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In: 15th annual conference of the international speech communication association
39. Pham V, Bluche T, Kermorvant C, Louradour J (2013) Dropout improves recurrent neural networks for handwriting recognition. ArXiv preprint [ArXiv: 1312:4569](https://arxiv.org/abs/1312.4569)
40. Norbert B, Hans-Peter K, Ralf S, Bernhar S (1990) The R\*-tree: an efficient and robust access method for points and rectangles. *ACM SIGMOD Record* 19(2):220–231
41. Yu Z, Xing X, Wei-Ying M (2010) Geolife: a collaborative social networking service among user, location and trajectory. *IEEE Data Eng Bull* 33(2):32–39
42. OpenStreetMap. <https://www.openstreetmap.org/>. Accessed on April 2019
43. Krumm J, Letchner J, Horvitz E (2007) Map matching with travel time constraints. SAE Technical Paper Series (2007)
44. Pattanayak S (2017) Introduction to deep-learning concepts and TensorFlow. Pro Deep Learning with TensorFlow, pp 89–152
45. Abdalla M, Hendawi A, Mokhtar HMO, Elgamal N, Krumm J, Ali M (2018) SimilarMove: similarity-based prediction for moving object future path. In: Proceedings of the 2nd ACM SIGSPATIAL workshop on prediction of human mobility, pp 15–24
46. Abdalla M, Hendawi A, Mokhtar HMO, Elgamal N, Krumm J, Ali M (2020) DeepMotions: a deep learning system for path prediction using similar motions. *IEEE Access*, 8
47. Tao Y, Faloutsos C, Papadias D, Liu B (2004) Prediction and indexing of moving objects with unknown motion patterns. In: Proceedings of the 2004 ACM SIGMOD international conference on management of data—SIGMOD 04

48. Zhou M, Tung AKH, Wu W, Ng WS (2013) A “semi-lazy” approach to probabilistic path prediction. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining—KDD 13
49. Cui J, Zhou X, Zhu Y, Shen Y (2018) A road-aware neural network for multi-step vehicle trajectory prediction. In: Database systems for advanced applications lecture notes in computer science, pp 701–716

# Hepatitis C Prediction Applying Different ML Classification Algorithms



**Md. Boktiar Hossain, Khandoker Hoque, Mohammad Atikur Rahman, Priya Podder, and Deepak Gupta**

**Abstract** Hepatitis C is a liver inflammation contracted from the Hepatitis C Virus (HCV). This disease is characterized by symptoms that appear relatively late into the course of the disease, thus, the diagnosis at an early stage is very challenging. In short, the efficiency of the prediction before permanent liver damage will annually save several patients. This work focuses on the following core purpose: To apply different machine learning algorithms for classifying this disease using cost-efficient tests to identify the disease's initial stages for early diagnosis and treatment of patients. Seven machine learning algorithms have been used in this study; logistic regression, K-nearest neighbors, decision tree, Random Forest, support vector machine, Naive Bayes, and LGBM on HCV datasets collected from the UCI repository. We evaluated the performance of those ML models with or without applying the SMOTE algorithm. In comparing these techniques from the perspective of confusion matrix, precision, recall, F1 score, accuracy, ROC, and AUC, it is found that LGBM provides 94.61% and 93.50% accuracy with and without SMOTE, respectively.

**Keywords** Accuracy · Decision tree (DT) · K-nearest neighbors (KNN) · Logistic regression (LR) · Precision · Random forest (RF) · Recall

---

Md. B. Hossain (✉) · K. Hoque · M. A. Rahman

School of Engineering, San Francisco Bay University, Fremont, CA, USA

e-mail: [m.boktiar.hossain@gmail.com](mailto:m.boktiar.hossain@gmail.com)

K. Hoque

e-mail: [khoque43977@student.sfbu.edu](mailto:khoque43977@student.sfbu.edu)

M. A. Rahman

Institute of Information and Communication Technology, BUET, Dhaka, Bangladesh

P. Podder

Dhaka National Medical College, Dhaka, Bangladesh

D. Gupta

Department of Computer Science and Engineering, Maharaja Agrasen Institute of Technology, Delhi, India

e-mail: [deepakgupta@mait.ac.in](mailto:deepakgupta@mait.ac.in)

## 1 Introduction

Hepatitis C can be described as one form of Hepatitis that results from the Hepatitis C Virus (HCV) [1]. This virus can be one of the main causes of liver cancer. A person can be affected by chronic Hepatitis due to this virus, which can lead him or her to the acute Hepatitis. As a result, one can suffer from severe illness of several weeks to a long debilitating illness right up to death [1, 2]. Therefore, HCV can be considered as a sort of a blood-borne virus. The forms of infection which are stated above are, to a large extent, used with blood portions that are comparatively small [3]. Some of the ways through which it can spread include; sharing drugs and needles used in injecting drugs, unsafe medical practices; receiving contaminated blood or blood plasma, unsafe practices of using injections although rare, promiscuous sexual behaviors [4–6].

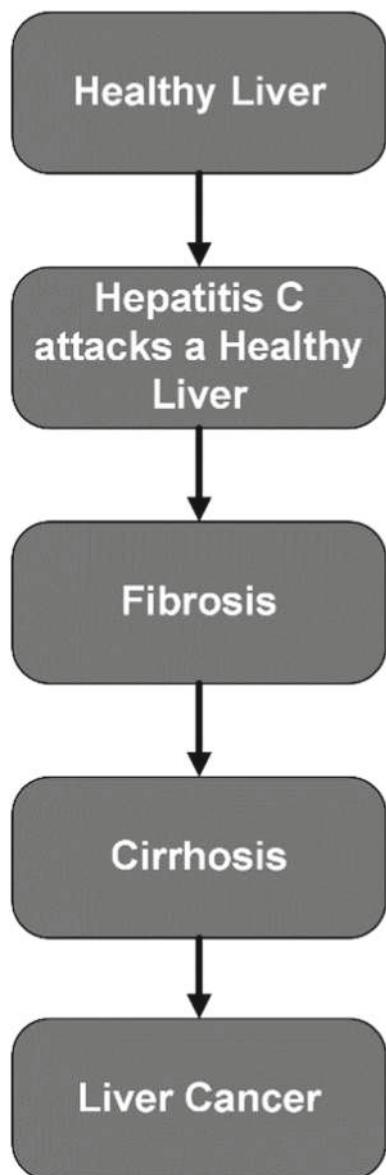
Using antiviral infections, HCV infection is treatable and within 95 percent in the cases cured, thus lowering the death rates arising from liver cancer or cirrhosis complications. Yet a very limited number of people receives treatment and diagnosis [2]. Hepatitis C affects majority of the people with few revealing symptoms in their initial stages. However, some symptoms such as fever, fatigue, jaundice, and headache should be observed after the time of two weeks up to six months after the HCV has entered the blood stream.

Chronic Hepatitis C is an acute illness that is still in its early stages, and it does not have a cure called a vaccine. This disease often triggers the emergence of severe infections in the body, such as hepatic cirrhosis, fibrosis, and cancer [7]. The different stages of HCV infection are shown in Fig. 1.

It is significant to highlight that HCV has many stages in the human body. Liver fibrosis normally develops in any response to tissue injury and damaged tissue. Similar cirrhosis is a Hepatic Fibrosis phase at the highest level with the implementation of hepatic architectonics and vasculature [8]. Liver cancer is likely to occur where an appropriate diagnosis is not well addressed. Hepatitis, detected and diagnosed correctly through blood samples also called liver tests together with the right medication may cure the disease. This liver test consists of two major serum biochemical enzymes referred to as alanine transferase or ALT and serum glutamic-oxaloacetic transaminase or SGOT. Where a patient has a higher level of ALT, the individual is more vulnerable to being infected with the Hepatitis virus. Recommend HCV test for the patient. The level of Hepatitis C is diagnosed at certain ranks of HCV at 12 weeks. Serological markers in blood serum as a tool for the prognosis of disease conditions and decreasing the burden on the health care system [9]. Acute Hepatitis or short-term phase Hepatitis is the first six months of HCV infections, after six months it turns into a chronic phase causing long-term HCV-associated illness [4].

Diagnosis of HCV is done in two stages. The first step is mainly concerned with the right diagnosis parameters while the second step is perhaps advocating for correct analysis of data.

**Fig. 1** Stages of HCV infection



Automation in deliveries of healthcare is expanding each and every day as many people embrace the use of artificial intelligence. According to one peculiar source, interest in AI and medicine projects prevailed over the interest in many other projects in the world economy at the end of recent years [5]. AI in medicine concerns the employment of automatism in diagnosing diseases as well as in supervising people who require medical assistance [6]. Introducing AI in an element of the activity of

prescribing medicine would enable it to absorb a considerable part of the work and relieve the medical professionals of the monotonous and time-consuming task that cannot be easily delegated to a machine [5]. SMOTE-based feature selection method has been applied in this paper, which not only helps to select the crucial features from the HCV dataset but also helps to overcome the problem of data imbalance. Therefore, the chance of improving the overall performance of the model will be increased.

The related work is depicted in the sub-section entitled “Literature Review.” The system proposed is explained in the “Methodology” section of the paper. The parts related to the outcome of the experiment and their analysis are called “Results and Discussion.” The final evaluation along with the plan for further studies is presented in “Conclusion.”

## 2 Literature Review

Ripa et al. [12] conducted a study using machine learning techniques to develop models for classifying HCV infections, specifically to predict the virus causing the illness. The research [12] incorporated a range of machine learning algorithms, such as Random Forest, CatBoost, Bagging Classifier, SGD Classifier, various Naive Bayes models (Gaussian, Bernoulli, and Multinomial), Linear Discriminant Analysis, Artificial Neural Networks (ANNs), and Multi-Layer Perceptron (MLP). Before training the models, the dataset was enhanced through preprocessing steps like normalization, filtering, and the application of SMOTE to balance the data [12].

Hashem et al. [13] established predictive models for the advanced liver fibrosis for CHCV patient using PSO, ADT, MReg, and GA. From these, ACC achieved the highest accuracy of 84 percent. 4%. The strength of an ADT is that can makes a difficult decision more accurate, flexible, but it may be slower than other methods and not as easily understood. Metwally et al. [14] suggested an ANN model for Hepatitis diagnosis, the model has an overall accuracy of 98. 44%. The ANN models are good for capturing non-linear relationships in the data but they are computationally expensive, and they are generally considered as “black boxes” in the sense that it is challenging to distill out how a decision was made by the model. Ma et al. [15] also used the extreme gradient boosting (XGBoost) algorithm in making the prognosis of HCV progression with an accuracy of 91%. 56%. One of the approaches that are well-optimized for large datasets and perform well in terms of efficiency is XGBoost but it is sensitive to overfitting if used with out-of-box parameters. Ahammed et al. [16] employed SMOTE for balancing their dataset and evaluated nine different classifiers on HCV prediction of which KNN achieved an accuracy of 94%. 40%. KNN is simple to implement and interpretable but it is may be computationally costly and influenced by noisy attributes. Feature selection and a DT classifier were employed by Ayeldeen et al. [17] in the prognosis of various phases of HCV fibrosis with final accuracy of 93%. 7%. Decision trees, we see, are easy to build and easy to interpret

but they suffer from being highly sensitive and, thus, often suffer from overfitting and also they may not be good with high dimensional data.

Nandipati et al. [18] applied ML models in Python for the HCV prediction using the Egyptian patient's dataset. The best accuracy of 54.56% was obtained by Random Forest in binary class, while KNN with R obtained 51. 06% accuracy in multiclass labels [18]. RF, APRI, and FIB-4 tests were applied in [19] for predicting and staging of Hepatic Fibrosis in Egyptian children. APRI achieved AUCs of 78%, 81.6%, and 77% for predicting any fibrosis, advanced fibrosis, and distinguishing between mild and advanced fibrosis, respectively [19]. FIB-4 showed similar performance with AUCs of 74%, 82.8%, and 78% for the same categories [19]. RF outperformed both APRI and FIB-4, with AUCs of 0.903, 0.894, and 0.822. [19]. The dataset owned by Li et al. [20], with the patients' history of 920 patients, was used for the creation of RFC, DTC, SVC, and LRC to evaluate the severity of liver fibrosis. RFC gave highest precision of 83% in [20].

All the previously presented approaches have a few of the weaknesses such as low accuracy, experiments with fewer parameters, experiments with fewer samples.

### 3 Methodology

In this method, we have used a dataset [22, 23] taken from the UCI repository. This has 615 instances and 12 features. The attributes are categorized in Table 1.

Jupyter notebook is used to run the experiment. PC configuration is 12 generation core i7 CPU, 16 GB RAM. Figure 2 shows the number of patients in each category, i.e., Hepatitis, Cirrhosis, Fibrosis, Blood Donor.

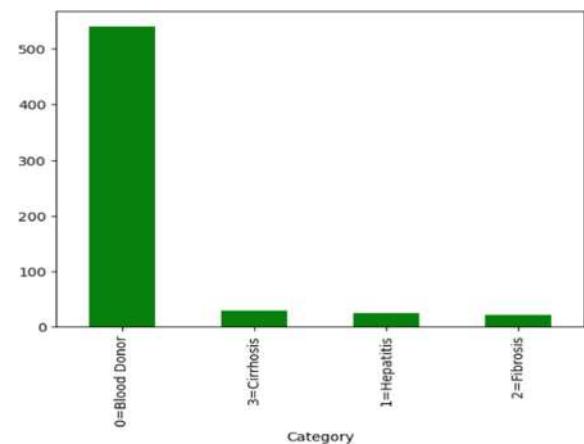
The flowchart of the data processing and machine learning pipeline for Hepatitis C data set is presented in Fig. 4. Data labeling step consists of qualifying or likely encoding the data, which probably means converting the raw data to a format that can be analyzed. Null value handling entails handling of missing values in the data set and can be handled by either imputing or removing the statistics. MinMaxScaler() operation is done using the MinMaxScaler, which scales the features to a prescribed range, often [0,1]. To determine the interactions between features, correlation heatmap, is produced, which is illustrated in Fig. 3. Class imbalance is handled in this case using Synthetic Minority Over-sampling Technique (SMOTE) that makes synthetic versions of samples on the minority classes.

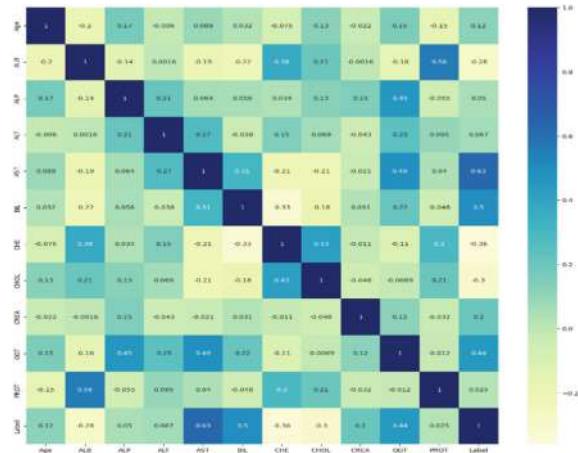
SMOTE is a sampling technique. From the neighbors of the pattern in the minority class, it randomly generates more instances of the said minority class. Real minority class samples are built with the help of features taken from the original data to reconstruct the above-mentioned individuals. To handle issues related to data imbalance, our proposed work applies the SMOTE approach. Equation (1) is used in SMOTE for creating a new minority class [21].

$$P_{newf} = (P_i + (P_{selectedF} - P_i) * t \quad (1)$$

**Table 1** Attributes of Hepatitis C dataset

S. No	Attributes	
1	X (Patient ID/No.)	
2	Category (diagnosis)	
	0 s	Suspect blood donor
	1	Hepatitis
	2	Fibrosis
	3	Cirrhosis
3	Age (in years) (20 years to 65 years range)	
4	Sex (f = female, m = male)	
5	ALB (Albumin in the blood)	
6	ALP (Alkaline phosphatase)	
7	ALT (Alanine amino-transferase) (status of the liver damage)	
8	AST (Aspartate amino-transferase in the liver)	
9	BIL (Bilirubin in the blood)	
10	CHE (Choline esterase-liver function)	
11	CHOL (Cholesterol)	
12	CREA (Creatinine blood test)	
13	GGT ( $\gamma$ -glutamyl-transferase)	
14	PROT (Protein blood test)	

**Fig. 2** Number of patients in each category

**Fig. 3** Correlation heatmap

A SMOTE first identifies  $P_i$ , set of feature and afterward it identifies the neighboring points in order to check the imbalance in the data [21]. It then finds the difference between the new feature set and the old one, multiplies that by a number between zero and one [21]. Lastly, it appends the outcome to the featureset to arrive at a new point on a definite line segment. These steps are followed for all the feature sets [25].

Here, 20% data is taken as test data and the remaining 80% is taken as training data and to make the testing random the random state is fixed as 42. Equations (1)–(6) describe precision, recall, accuracy, FNR, TNR, and FPR.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

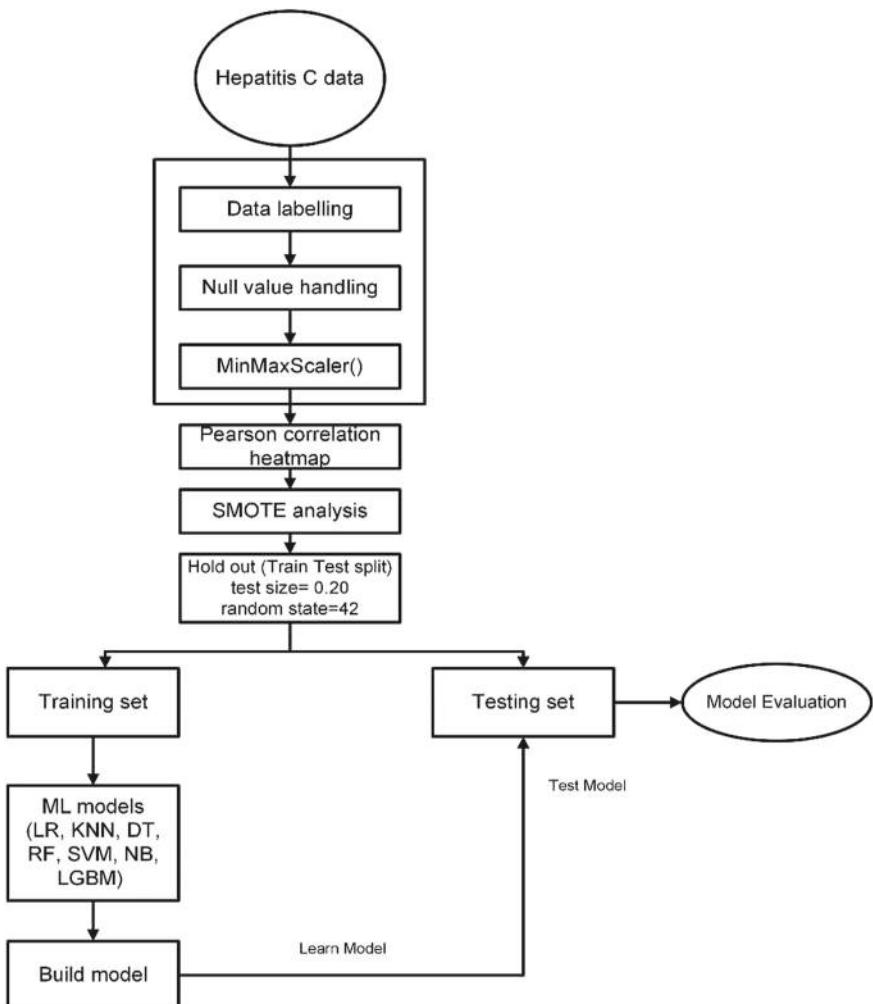
$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

$$\text{FNR} = \frac{\text{FN}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{TNR} = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (5)$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (6)$$

An area under the receiver operating characteristic (ROC) curve shows ratio of true positive rate (TPR) to the false positive rate (FPR) for each solution [23, 24],



**Fig. 4** Work flowchart

while true positive rate is calculated by the ratio of true positive across the total positive, false positive rate is computed as the ratio of false positive to the total number of actual negative cases [24]. AUC [24], is the measure of the classifier's ability in terms of the separability of its classes. It follows that a higher AUC have a better ability to classify the positives from negatives [24].

## 4 Results and Discussion

Tables 2, 3, 4 show the classification metrics and accuracy for KNN, LR, and DT, both without and with SMOTE. Without SMOTE, KNN, LR, and DT, perform well for Class 0 and Class 1. KNN and LR struggle with Class 2 detection, whereas DT shows better performance. SMOTE helps in slightly improving detection of minority classes (Class 2 and Class 3) but not significantly. KNN and LR show some improvement in Class 2, but still low precision and recall. DT shows the best improvement with SMOTE, especially for Class 2 and Class 3. Overall accuracy decreases slightly with SMOTE for KNN and LR but improves for DT.

Table 9 shows the Average TPR, FPR, TNR, and FDR with and without SMOTE implementation. LR incorporates a satisfactory TPR of 0.72 and FPR of 0.04; thus, the model distinguishes between the two classes well, with minimal false positives.

**Table 2** Classification report of KNN

Class	Precision	Recall	F1-score	Accuracy
<i>Without SMOTE</i>				
0	0.93	1.00	0.96	0.9187
1	1.00	0.00	0.00	
2	0.00	0.00	0.00	
3	1.00	0.83	0.91	
<i>With SMOTE</i>				
0	0.96	0.96	0.96	0.9024
1	0.40	0.40	0.40	
2	0.17	0.25	0.20	
3	1.00	0.67	0.80	

**Table 3** Classification report of LR

Class	Precision	Recall	F1-score	Accuracy
<i>Without SMOTE</i>				
0	0.92	1.00	0.96	0.9189
1	0.00	0.00	0.00	
2	0.00	0.00	0.00	
3	1.00	0.83	0.91	
<i>With SMOTE</i>				
0	0.98	0.87	0.92	0.8458
1	0.10	0.20	0.13	
2	0.33	0.50	0.40	
3	0.55	1.00	0.71	

**Table 4** Classification report of DT

Class	Precision	Recall	F1-score	Accuracy
<i>Without SMOTE</i>				
0	0.99	0.94	0.97	0.9024
1	0.00	0.00	0.00	
2	0.40	1.00	0.57	
3	0.71	0.83	0.77	
<i>With SMOTE</i>				
0	0.99	0.98	0.99	0.9268
1	0.30	0.60	0.40	
2	0.50	0.25	0.33	
3	1.00	0.67	0.80	

**Table 5** Classification report of RF

Class	Precision	Recall	F1-score	Accuracy
<i>Without SMOTE</i>				
0	0.98	1.00	0.99	0.9389
1	0.25	0.20	0.22	
2	0.00	0.00	0.00	
3	0.86	1.00	0.92	
<i>With SMOTE</i>				
0	0.98	1.00	0.99	0.9430
1	0.60	0.60	0.60	
2	0.25	0.25	0.25	
3	1.00	0.67	0.80	

**Table 6** Classification report of SVM

Class	Precision	Recall	F1-score	Accuracy
<i>Without SMOTE</i>				
0	0.92	1.00	0.96	0.9186
1	0.00	0.00	0.00	
2	0.00	0.00	0.00	
3	1.00	0.83	0.91	
<i>With SMOTE</i>				
0	0.98	0.89	0.93	0.8780
1	0.12	0.20	0.15	
2	0.25	0.50	0.33	
3	0.67	1.00	0.80	

**Table 7** Classification report of NB

Class	Precision	Recall	F1-score	Accuracy
<i>Without SMOTE</i>				
0	0.96	0.94	0.95	0.8862
1	0.40	0.40	0.40	
2	0.25	0.25	0.25	
3	0.56	0.83	0.67	
<i>With SMOTE</i>				
0	0.98	0.91	0.94	0.8780
1	0.22	0.40	0.29	
2	0.25	0.25	0.25	
3	0.60	1.00	0.75	

**Table 8** Classification report of LGBM

Class	Precision	Recall	F1-score	Accuracy
<i>Without SMOTE</i>				
0	0.99	1.00	1.00	0.9461
1	0.50	0.20	0.29	
2	0.33	0.25	0.29	
3	0.67	1.00	0.80	
<i>With SMOTE</i>				
0	0.98	0.99	0.99	0.9350
1	0.50	0.60	0.55	
2	0.25	0.25	0.25	
3	1.00	0.67	0.80	

KNN has a higher TPR (0.76) but a higher FPR (0.10), while the proposed shows a lower TPR (0.60) yet a lower FPR (0.05). DT presents a lower TPR (0.47) than the proposed method but a slightly higher FDR (0.55). RF offers a good performance with a TPR ranging between 0 percent. 60 and low FDR at 0.42, thus, meaning a low rate of false positives. SVM gives a moderate TPR of 0.63 and a relatively low FPR of 0.05. NB presents high TPR of 0.68 and the lowest FPR of 0.03, thus, depicting high capability in true positive identification. LightGBM (LGBM) generates the peak TNR of 0.98 with a TPR of 0.64 and the lowest FDR of 0.

without SMOTE, LR shows reduction of TPR (0.53) while FPR is elevated to 0.12 which clearly suggests that the number of samples is misclassified compared to the best result obtained when SMOTE is used. K-nearest neighbors (KNN) has ability of having a lower TPR equal to 0.44 and higher FPR equal to 0.14 when compared to other classifier when it is tried without SMOTE. RF presents good TPR (0.65) and FPR (0.05), and holds its ground even when SMOTE is not used. NB has

**Table 9** Average TPR, FPR, TNR, and FDR

Model	Average TPR	Average FPR	Average TNR	Average FDR
<i>With SMOTE</i>				
LR	0.72	0.04	0.96	0.47
KNN	0.60	0.05	0.95	0.51
DT	0.47	0.05	0.95	0.55
RF	0.60	0.04	0.96	0.42
SVM	0.63	0.05	0.95	0.47
NB	0.68	0.03	0.97	0.46
LGBM	0.64	0.02	0.98	0.40
<i>Without SMOTE</i>				
LR	0.53	0.12	0.88	0.19
KNN	0.44	0.14	0.89	0.29
DT	0.62	0.05	0.95	0.10
RF	0.70	0.03	0.97	0.31
SVM	0.53	0.12	0.88	0.12
NB	0.64	0.04	0.96	0.48
LGBM	0.61	0.03	0.97	0.37

**Table 10** Comparative analysis

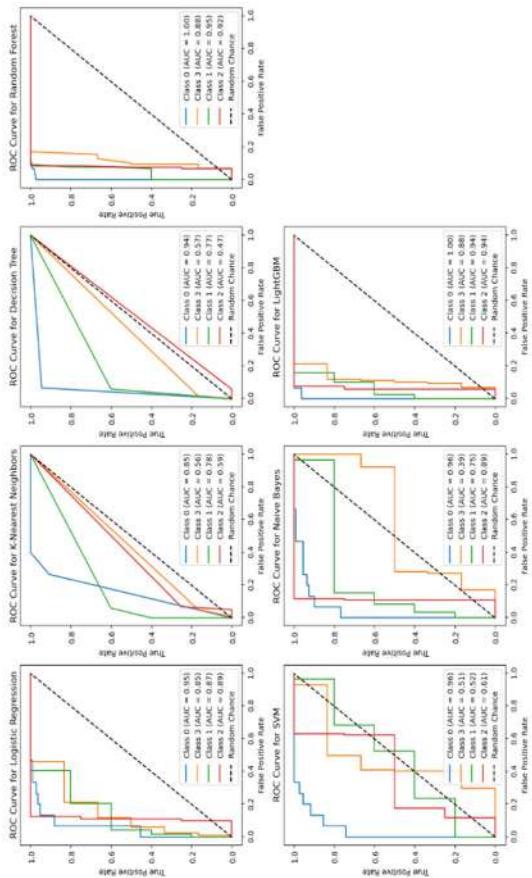
References	Model	Recall (%)	Precision (%)	F1-score (%)	Accuracy (%)
[7]	SVM	31.26	35.68	35.89	33.47
	RF	36.55	38.75	37.99	37.52
	DT	35.10	35.34	34.79	35.45
	BGLM	31.25	30.75	30.45	32.23
	HPM	40.56	41.22	42.33	41.54
[26]	KNN	64	78	—	89.43
	NB	72	69	—	90.24
	RF	79	83	—	94.31
[27]	NB	—	—	—	91.89
	KNN	—	—	—	93.09
	DT	—	—	—	93.09
[28]	DT	—	—	—	88
	LR	—	—	—	91
	SVM	—	-	—	92
Proposed method	LGBM (Without SMOTE)	62.25	61.25	59.5	94.61
	LGBM (With SMOTE)	68.29	62.69	58.25	93.50

a good TPR of 0. 64 and an FPR of 0. 04, hence a good performance without the use of SMOTE.

LightGBM (LGBM) attains average TPR, FPR, TNR, FDR of 0.61, 0. 03, 0.97, and 0.37 without SMOTE (Fig. 5).

## 5 Conclusion

In this work, seven machine learning algorithms have been used in this study; LR, KNN, DT, RF, SVM, NB, and LGBM on HCV datasets collected from UCI repository. However, this work has some weaknesses, namely, the utilization of only a small number of datasets, the absence of clinical data, and no controlled experiment. Our future work includes the inclusion of other features concerning Hepatitis C aiming toward the construction of more successful and effective machine learning algorithms. In addition, it is suggested that a clinical trial should be designed to evaluate the efficiency of the proposed methods in practice. To sum up, our work offers a positive outlook on the algorithm-based diagnosis of the Hepatitis C disease at its beginning stages, which will likely enhance patients' quality of life and reduce mortality rates.



**Fig. 5** ROC curves

## References

1. Manns MP, Buti M, Gane E et al (2017) Hepatitis C virus infection. *Nat Rev Dis Primers* . 3. <https://doi.org/10.1038/nrdp.2017.6.17006>
2. Who WHO (2020) Hepatitis C. <https://www.who.int/news-room/fact-sheets/detail/hepatitis-c>
3. Pietschmann T, Brown RJP (2019) Hepatitis C virus. *Trends Microbiol* 27(4):379–380. <https://doi.org/10.1016/j.tim.2019.01.001>
4. Prediction of fibrosis progression rate in patients with chronic hepatitis C genotype 4: role of cirrhosis risk score and host factors. *J Interferon Cytokine Res.* 37(3):97–102. <https://doi.org/10.1089/jir.2016.0111>
5. Ahuja AS (2019) The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ.* 7 <https://doi.org/10.7717/peerj.7702.e7702>
6. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K (2019) The practical implementation of artificial intelligence technologies in medicine. *Nat Med* 25(1):30–36. <https://doi.org/10.1038/s41591-018-0307-0>
7. Lilhore UK, Manoharan P, Sandhu JK et al (2023) Hybrid model for precise hepatitis-C classification using improved random forest and SVM method. *Sci Rep* 13:12473. <https://doi.org/10.1038/s41598-023-36605-3>
8. Sharma A, Arora A, Gupta A, Singh PK (2022) Data-centric approach to hepatitis C virus severity prediction. In: Abraham A, Gandhi N, Hanne T, Hong TP, Nogueira Rios T, Ding W (eds) Intelligent systems design and applications. ISDA 2021. Lecture Notes in Networks and Systems, vol 418. Springer, Cham. [https://doi.org/10.1007/978-3-030-96308-8\\_39](https://doi.org/10.1007/978-3-030-96308-8_39)
9. Elgarably A et al (2016) Hepatitis C in Egypt—past, present, and future. *Int J Gen Med* 10:1–6. <https://doi.org/10.2147/IJGM.S119301>
10. Rahman F, Das D, Sami A, Podder P, Michael DL (2024) Liver cirrhosis prediction using logistic regression, naïve bayes and KNN. *Int J Sci Res Arch* 12(01):2411–2420
11. Hoque R, Billah M, Debnath A, Hossain SS, Sharif NB (2024) Heart disease prediction using SVM. *Int J Sci Res Arch* 11(2):412–420
12. Ripa R, Uddin KMM, Alam MJ et al (2024) Hepatitis C prediction using machine learning and deep learning-based hybrid approach with biomarker and clinical data. *Biomed Mater Dev.* <https://doi.org/10.1007/s44174-024-00197-x>
13. Hashem S et al (2017) Comparison of machine learning approaches for prediction of advanced liver fibrosis in chronic hepatitis C patients. *IEEE/ACM Trans Comput Biol Bioinf* 15(3):861–868
14. Metwally NF, AbuSharekh EK, Abu-Naser SS (2018) Diagnosis of hepatitis virus using artificial neural network. *Int J Acad Pedagog Res (IJAPR)* 2(11):1–8
15. Ma L, Yang Y, Ge X, Wan Y, Sang X (2020) Prediction of disease progression of chronic hepatitis C based on XGBoost algorithm. In: 2020 International conference on robots and intelligent system (ICRIS). IEEE, pp 598–601
16. Ahammed K, Satu MS, Khan MI, Whaiduzzaman M (2020) Predicting infectious state of hepatitis c virus affected patient's applying machine learning methods. In: 2020 IEEE Region 10 Symposium (TENSYMP). IEEE, pp 1371–1374
17. Ayeldeen H, Shaker O, Ayeldeen G, Anwar KM (2015) Prediction of liver fibrosis stages by machine learning model: a decision tree approach. In: 2015 Third World conference on complex systems (WCCS). IEEE, pp 1–6
18. Nandipati SC, XinYing C, Wah KK (2020) Hepatitis C virus (HCV) prediction by machine learning techniques. *Appl Modelling Simul* 4:89–100
19. Barakat NH, Barakat SH, Ahmed N (2019) Prediction and staging of hepatic fibrosis in children with hepatitis c virus: a machine learning approach. *Healthcare Inf Res* 25(3):173–181. <https://doi.org/10.4258/hir.2019.25.3.173>
20. Li N, Zhang J, Wang S, Jiang Y, Ma J, Ma J, Dong L, Gong G (2019). Machine learning assessment for severity of liver fibrosis for chronic HBV based on physical layer with serum markers. *IEEE Access*, 7:124351-124365

21. Samreen S (2024) Accurate Prediction of Stage of Hepatitis C Virus Through a Stacking Ensemble. In: Nanda SJ, Yadav RP, Gandomi AH, Saraswat M (eds) Data science and applications. ICDSA 2023. Lecture Notes in Networks and Systems, vol 821. Springer, Singapore. [https://doi.org/10.1007/978-981-99-7814-4\\_38](https://doi.org/10.1007/978-981-99-7814-4_38)
22. <https://archive.ics.uci.edu/dataset/571/hcv+data>
23. Hoffmann G, Bietenbeck A, Lichtenhagen R, Klawonn F (2018) Using machine learning techniques to generate laboratory diagnostic pathways—a case study. *J Lab Prec Med* 3(6)
24. Bharati S, Rahman MA, Podder P (2018) Breast cancer prediction applying different classification algorithm with comparative analysis using WEKA. In: 2018 4th international conference on electrical engineering and information and communication technology (iCEEiCT). IEEE, pp 581–584
25. Safdari R, Deghatipour A, Gholamzadeh M, Maghooli K (2022) Applying data mining techniques to classify patients with suspected hepatitis C virus infection. *Intell Med* 2(04):193–198
26. Syafaâ L, Zulfatman Z, Pakaya I, Lestandy M (2021) Comparison of machine learning classification methods in hepatitis C virus. *J Online Informatika* 6(1):73–78
27. Islam S, Rehman AU, Javaid S, Ali TM, Nawaz A (2022) An integrated machine learning framework for classification of cirrhosis, fibrosis, and hepatitis. In: 2022 third international conference on latest trends in electrical engineering and computing technologies (INTELLECT), Karachi, Pakistan, pp 1–6. <https://doi.org/10.1109/INTELLECT55495.2022.9969404>
28. Ara A, Sami A, Michael DL, Bazgir E, Mandal P (2024) Hepatitis C prediction using SVM, logistic regression and decision tree. *World J Adv Res Rev* 22(2):926–936

# Alzheimer's Disease Detection Using Hybrid Radial Basis Function Neural Network Integrated with Harris Hawk Optimization



Nair Bini Balakrishnan<sup>ID</sup>, Anitha S. Pillai, and Jisha Jose Panackal

**Abstract** Alzheimer's disease (AD) is a serious neurological brain illness that results in the death of brain cells and the loss of mental and memory abilities in sufferers. Planning an appropriate course of treatment for AD requires early and accurate identification of the disease. We presented a novel approach to AD detection in this paper by combining the benefits of meta-heuristic optimization with deep learning. Hence, the Radial Basis Function Neural Network (RBFNN) and the Harris Hawks Optimization (HHO) algorithm are introduced to accurately detect AD using brain MRI data. The brain MRI images are acquired from the dataset, and then preprocessed to enhance quality in preparation for additional analysis. The RBFNN was trained using the preprocessed database to ascertain the pattern difference between normal and AD images. As a result, the HHO is used to tune the RBFNN hyperparameters to reduce processing requirements and increasing classification accuracy. The Python implementation of the proposed framework was verified using Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Open Access Series of Imaging Studies (OASIS). The results of the experiment demonstrated that the proposed strategy achieved very good performance on both datasets. The proposed method produced results for the ADNI dataset with 99.56% accuracy, 99.72% precision, 99.34% recall, and 99.51% f-measure. Similarly, the OASIS dataset produced results with accuracy, precision, recall, and f-measure of 99.27%, 99.05%, and 99.11%, respectively. The experimental and comparative analysis showed that the proposed technique outperformed the conventional models in terms of accuracy, precision, recall, and f-measure.

---

N. B. Balakrishnan (✉) · A. S. Pillai

Department of Computer Applications, Hindustan Institute of Technology and Science, Chennai,  
India

e-mail: [rp.21703012@student.hindustanuniv.ac.in](mailto:rp.21703012@student.hindustanuniv.ac.in); [biniremesh@gmail.com](mailto:biniremesh@gmail.com)

A. S. Pillai

e-mail: [anithasp@hindustanuniv.ac.in](mailto:anithasp@hindustanuniv.ac.in)

J. J. Panackal

Department of Computer Science, Sacred Heart College, Thrissur, Kerala, India

**Keywords** Alzheimer's disease · Deep learning · Harris Hawks Optimization · Radial Basis Function Neural Network · Disease diagnosis

## 1 Introduction

Alzheimer's disease is a neurological disorder that affects people in different ways. Mental illness, confusion, memory loss, and issues with speaking, writing, and reading are common in AD patients [1]. This medical condition may also cause the person to lose memories of their family and way of life [2]. Three phases usually characterize the progression of AD: Moderate, mild, and extremely mild. It is challenging to diagnose this medical condition until it reaches its middle stage [3]. However, early detection is essential for preventing damage to brain tissue and ensuring adequate treatment planning [4]. In addition, the standard diagnosis of AD involves the use of the Mini-Mental State Examination (MMSE), patient demographics and medical history, physical and neurological tests among other techniques [5]. This AD detection method is often time-consuming and prone to human error. Owing to these issues with the traditional approach, an automated and reliable framework for AD diagnosis is needed [6]. Medical experts investigate the neurological functioning of the human brain through Computed Tomography (CT), MRI, and X-ray scanning [7]. By examining these medical images, healthcare experts can determine the stages of AD with accuracy. Moreover, MRI data outperformed the other imaging modalities in terms of accuracy in AD classification and detection [8]. Medical researchers have been working on developing computed-aided detection frameworks for accurate and timely disease diagnosis in the last several years [9]. These frameworks forecast the different stages of AD by utilizing artificial intelligence techniques like deep learning (DL) and machine learning (ML) [10].

MRI data is used to train the ML and DL algorithms to identify the features that set the normal and AD stages. In addition, the supervised or unsupervised learning techniques are used to train these models [11]. In supervised learning, the model is trained on labeled data (the database includes labels for both normal and distinct AD phases), whereas unsupervised learning uses unlabeled data to train the model [12].

The recognition of AD patients using MRI images produces better accuracy when compared to existing works [13]. Consequently, ML and DL algorithms have been developed for accurate and timely identification of AD cases [14]. Earlier studies used a variety of machine learning models, including decision tree approach, k-nearest neighbor (KNN), gradient boosting algorithm (GBA), support vector machine (SVM), and random forest classifier (RF) [15].

Despite the fact that these models were more accurate than manual exams, they had drawbacks such as longer training times, poor generalization, and overfitting. [16]. Conversely, research employing deep learning models made use of methods like Elman neural networks (ENN), deep belief networks (DBNs), convolutional neural networks (CNN), deep neural networks (DNNs), etc. [17]. Compared to ML models, these models provided somewhat higher accuracy and generalizability [18].

As a result, deep learning models are frequently used for various medical picture disease classification and detection [19]. Although, deep learning models still facing number of problems, such as the necessary for large datasets, high processing costs, and overfitting [20]. Hybrid models, which incorporate the best features of several methodologies are introduced to increase efficacy and performance to solve these difficulties.

An innovative method is introduced to tackle these issues by combining the HHO for Alzheimer's disease identification with a Radial Basis Function Neural Network (RBFNN). The major contributions of the presented study are described below:

- The proposed work developed a novel AD classification strategy leveraging the efficiency of deep learning and meta-heuristic optimization approaches.
- The proposed algorithm's Radial Basis Function Neural Network (RBFNN) effectively learns and accurately understands the pattern difference between normal and AD instances through intensive training with preprocessed data.
- The proposed strategy's Harris Hawk Optimization refines the RBFNN hyperparameters by fine-tuning them to their optimal range, reducing computational efforts, and enhancing the model's overall performance.
- The presented algorithm was validated across ADNI and OASIS databases, and the results are evaluated in terms of accuracy, precision, recall, and f-measure.

## 2 Related Works

The following describes a few recent research that are related to the planned work. One of the most prevalent types of dementia in developing nations, AD is the leading cause of death for elderly people. A study by Janghel and Rathore [21] used the ADNI database to identify AD. Identifying AD and normal from the ADNI dataset was done using a deep learning technique. During preparation, image conversion and scaling were first carried out to improve the quality of the database. Additionally, features were extracted and identified from the MRI images using CNN and VGG-16. Moreover, the SVM classifier was ultimately used to identify and differentiate the AD class from regular images. This methodology had an accuracy of 91.16%, according to the testing findings. Unfortunately, the processing power and training time of this model are very high.

Sathiyamoorthi et al. [22] developed a deep learning architecture-based computer-aided diagnosis framework for diagnosing AD accurately using brain MRI images. The noisy features in the MRI images were removed using a 2D Adaptive Bilateral Filter. Here, the images were acquired by scanning them with multiple devices. Consequently, the photos' contrast and brightness were improved by applying Adaptive Histogram Adjustment. Furthermore, a two-dimensional Gray Level Co-occurrence Matrix and Adaptive Mean Shift Modified Expectation Maximization were utilized for tasks such as feature selection and picture segmentation. Ultimately, from the normal photos, a Deep Convolutional Neural Network (DCNN) method was created to recognize and categorize AD cases. The outcomes of the implementation

clearly show that the developed method boosted accuracy and productivity. However, the system becomes more complex and less interpretable when several strategies are combined into a single framework.

A minimal recurrent neural network (minimal) approach was presented by Nguyen et al. [23] for the efficient diagnosis of AD using the ADNI database. The goal of this work is to address the scalability problem in AD detection and categorization. In order to address the missing value in the dataset and enhance database consistency, this study employed three distinct approaches. Long-term memory RNN architecture, which allows the system to learn sequentially and provides improved accuracy, was used in the developed study. Even though this method required less computational labor and produced higher accuracy, it is difficult to avoid overfitting in this work.

A hybrid method for identifying and categorizing Alzheimer's illnesses using MRI images was described by Dua et al. [24]. The main goal of this work is to solve the problems associated with the detection and classification of AD using traditional classification methods such as SVM, CNN. When analyzing huge databases, these models encounter difficulties. Using ensemble techniques to integrate CNN, RNN, and LSTM architectures, this study offered a dependable solution. According to the experimental data, the accuracy of the model increased from 89.75% to 92.22% when these models were combined into a single strategy. Nevertheless, concerns like interpretability and scalability have not been addressed by this methodology.

Habiba et al. [25] introduced a novel method that combined CNN's VGG 16 architecture with an ensemble classifier. This method uses an efficient analysis of brain MRI scans to identify and categorize AD classes from normal ones. This method obtained 97.09% accuracy when validated against the publicly available ADNI database. However, as the dataset gets bigger, its performance gets worse.

### 3 Problem Statement

AD is a neurological condition that results in memory loss, cognitive decline, and aberrant behavior. Early identification is essential for managing the condition and improving patient outcomes [14]. Traditional approaches of AD identification involve a time-consuming and expensive combination of clinical evaluations, neuroimaging, etc. Due to these issues, an automated diagnostic tool that rapidly diagnoses and treats AD using MRI data must be developed. The automatic detection of AD determined using deep learning models. These algorithms are able to differentiate between normal and AD cases based on specific characteristics and patterns by examining the MRI images as input. DL-based models comprise an overarching architecture that includes data collecting, data processing, and classification models. MRI images of the brain are obtained from hospitals or internet resources in the data collection phase. This database serves as the foundation for disease diagnosis. To enhance their uniformity and quality for further analysis, the raw images are filtered, denoised, and normalized during the preprocessing phase. In order to teach the DL approaches in the classification model to distinguish between the features of the normal and

AD classes, the filtered database is used. Although DL-based models outperformed classical models, they were not without limitations. Firstly, the training of the DL model uses large databases and a lot of processing power, which increases the model's implementation costs. Second, the generalization efficiency of DL techniques in real-world contexts is limited because they frequently become overtrained on the training sequence. Thirdly, the efficiency of processing big data volumes is compromised by the limitations of the current deep learning algorithms. The current methods are unsuitable for actual AD detection in clinical settings due to these problems. We suggested a cooperative framework that makes use of the effectiveness of deep learning and optimization algorithms with a natural theme to solve these problems.

## 4 Proposed Strategy for AD Detection and Classification

In order to diagnose Alzheimer's disease, this study suggested a novel classification technique that makes use of the efficiency of deep learning and meta-heuristic optimization algorithms. The suggested technique combines the advantages of the Harris Hawks Optimization with the Radial Basis Function Neural Network for precise brain MRI image-based Alzheimer's disease diagnosis. The proposed method uses brain MRI images that includes both healthy and Alzheimer's disease cases.

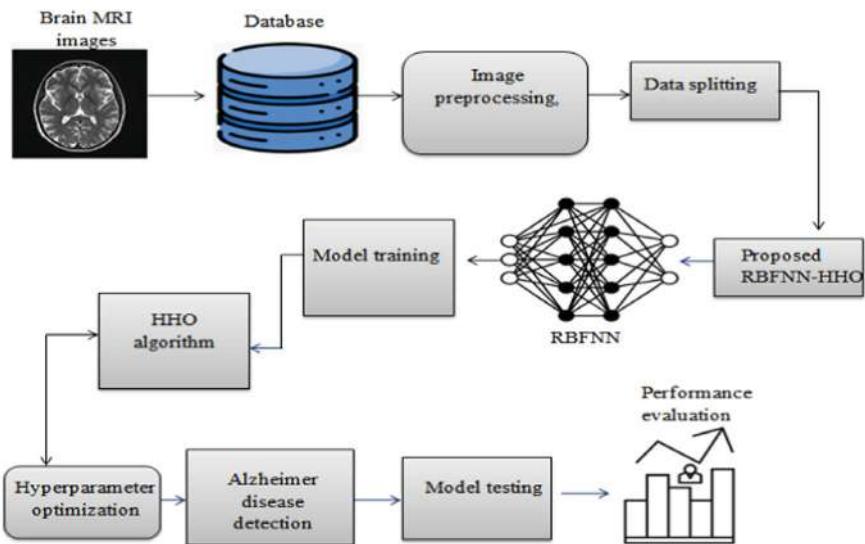
Preprocessing techniques are performed after data collection to enhance the quality of the images for further analysis. These preprocessed images are fed into the proposed RBFNN-WOA, which classifies and predicts AD. RBFNN is one of the deep learning techniques that recognizes the difference between AD and regular images. In addition, the RBFNN is trained to identify and classify the AD instances using the preprocessed database. The HHO algorithm then finds the optimal value for the RBFNN model parameters for enhancing both model training and classification accuracy. Figure 1 depicts the architecture of the proposed methodology.

### 4.1 Data Collection

Brain MRI images are first acquired in order to identify and categorize AD. Two publicly available datasets, like the Open Access Series of Imaging Studies (OASIS) and the Alzheimer's Disease Neuroimaging Initiative (ADNI) were employed.

#### ADNI Dataset

The ADNI dataset is a collection of MRI brain images that are publicly available for research on the diagnosis and classification of AD [26]. Each individual in this dataset has a wealth of information available on them, including recruitment procedures, demographic data, physical exams, and cognitive testing. Brain MRI samples from three classes—Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), and Common Normal (CN)—are included in the ADNI database. 819 brain MRI



**Fig. 1** Proposed strategy

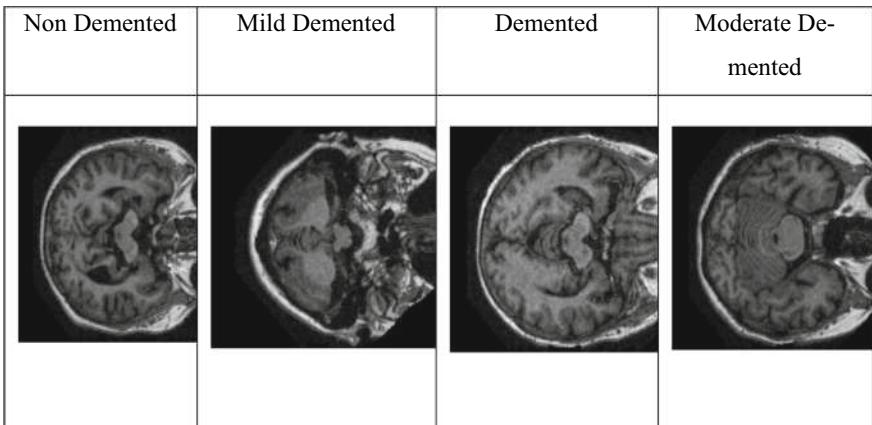
samples—229 CN pictures, 398 MCI images, and 192 AD images—are included in the database. Figure displays samples of the brain MRI scans that are stored in the ADNI database (Fig. 2).

### OASIS Dataset

80,000 brain MRI scans from the OASIS database are categorized into four different classes non-demented, very mildly demented, mildly demented, and moderately demented that correspond to the different stages of Alzheimer's disease progression [27]. To give a precise and comprehensive range of the illness stages, MRI pictures from 461 patients were gathered. The database is 1.3 GB in size, and the MRI scans are provided in.jpg format. Figure 3 shows the samples from the OASIS database.

**Fig. 2** Samples images from the ADNI dataset

AD	MCI	CN



**Fig. 3** Sample images from the OASIS dataset

## 4.2 Preprocessing

In order to accurately diagnose the condition, picture preprocessing which transforms the raw input photos into a standardized format is a crucial step in AD detection. Image scaling, normalization, denoising, contrast enhancement, segmentation, and data augmentation are some of the preprocessing procedures. Image resizing refers to bringing the input images' dimensions down to a standard size. By ensuring consistency across the databases, this stage facilitates the processing and learning of the images by categorization models. The pixel values were then scaled to fall between 0 and 1 using normalization. In this case, the resized photos were normalized using the z-score algorithm. This step speeds up the classifier convergence during training. Denoising was also done using Gaussian filtering, which smooths the input images and gets rid of noise features. Then, the histogram equalization was used to improve the contrast of the images.

### RBFNN for AD Detection

The RBFNN is a type of artificial neural network that can simulate complex interactions between inputs and outputs [28]. This neural network uses Radial Basis Functions (RBFs) as activation functions to recognize and approximate the complex patterns and relations found in the input data. Moreover, the RBFNN is used to interpret brain MRI samples in order to identify AD.

An RBFNN consists of input layer, the hidden layer, and the output layer. The input layer of the RBFNN feeds the preprocessed brain MRI samples into the hidden layer. Its core component is the hidden layer of the network, which processes input to differentiate between patterns in normal and AD. This RBF efficiently capture the correlations and patterns identified in the MRI brain data. The feature map is generated by highlighting the patterns, relationships that set apart normal and AD occurrences. The performance of the hidden layer is dependent on the type of RBF that the

network employs. RBFs come in various forms, including Inverse Multiquadratic, Multiquadratic, and Gaussian functions. In the paradigm that was described, the RBF was the Gaussian function, which can be stated mathematically in Eq. (1).

$$G(ip - ce_j) = e^{\left(-\frac{ip - ce_j^2}{2\lambda_j^2}\right)} \quad (1)$$

where  $\lambda$  represents the control variable,  $ce_j$  defines the center vector, and  $\|ip - ce_j\|$  denotes the distance between the center and the input vector. At each iteration, the Gaussian function of RBFNN adjusts its parameters such as width, center, weight, bias, and number of RBFs, enabling the system to identify the interconnections within the input images more effectively. Then, the learned feature map will be forwarded into the output layer of the system. The output layer performs the AD detection based on the learned features and patterns expressed in Eq. (2).

$$Op_i(ip) = \sum_{j=1}^k w_{gij} G(ip - ce_j) \quad (2)$$

where  $ip$  represents the input data,  $w_{gij}$  represents the weights interconnecting the  $j$ th hidden layer and  $i$ th output layer,  $k$  indicates the number of neurons present in the hidden layer, and  $Op_i$  denotes the prediction result. The prediction output of the output layer is passed through a sigmoid activation function, which provides probabilities that the input image is either normal or AD case. If the probability value is greater than 0.5, the system predicts the image as “AD,” or else the system identifies the image as “Normal.” The error prediction is obtained by determining the Mean Square Error, represented in Eq. (3).

$$\text{Loss} = \frac{1}{m} \sum_{i=1}^m \left( Op_i - \sum_{j=1}^n w_{gij} \exp(ip - ce_j^2) \right)^2 \quad (3)$$

where  $m$  denotes the number of training samples and  $n$  indicates the number of RBFs. In the training phase, this prediction error is reduced by optimally refining the RBFNN parameters like width, center, weight, bias, neuron count, etc. The proposed work performs the hyperparameter optimization of RBFNN using the HHO algorithm.

### 4.3 HHO for Hyperparameter Optimization

The chasing manner and cooperative nature of Harris Hawks served as the basis for the development of HHO, a nature-inspired optimization technique [29]. Mathematical models of the Harris Hawks’ distinct behavior are used to address a variety of practical optimization issues. In the produced work, the HHO algorithm was used to determine

the ideal RBFNN hyperparameter value within the search range. Setting a maximum iteration count for the Harris Hawk population is the first step in the hyperparameter optimization process utilizing the HHO algorithm. The hyperparameter sequence of RBFNN is indicated by each Harris Hawk throughout the population. Equation (4) presents the population's random initialization using HHO mathematically.

$$Hyp = \begin{bmatrix} hy_1^1 hy_1^2 \dots hy_1^x \dots hy_1^d \\ hy_2^1 hy_2^2 \dots hy_2^x \dots hy_2^d \\ \vdots \dots ; \dots ; \\ hy_X^1 hy_X^2 \dots hy_X^x \dots hy_X^d \end{bmatrix}, \begin{cases} x = 1, 2, 3, \dots, X \\ d = 1, 2, 3, \dots, D \end{cases} \quad (4)$$

where  $Hyp$  indicates the parameter population and  $X$  defines the population size. HHO's goal is to minimize the loss that the RBFNN model experiences by determining the hyperparameter sequences' ideal (best) value. Equation (5) presents the objective criterion of the HHO algorithm.

$$\text{Objective} = \underset{\text{minimize}}{\left( \frac{1}{m} \sum_{i=1}^m \left( \text{Op}_i - \sum_{j=1}^n w g_{ij} \exp(ip - ce_j^2) \right)^2 \right)} \quad (5)$$

Based on this objective function, the fitness was determined for each parameter sequence in the population. The fitness calculation is expressed in Eq. (6).

$$\text{Fitness}(hy) = \frac{1}{\text{Loss}} \quad (6)$$

where  $\text{Fitness}(hy)$  indicates the fitness value of the parameter sequence. The loss that the RBFNN model experiences for a given parameter is inversely related to the sequence's fitness. The fitness will be low if the model yields a significant loss for a given value of the parameter sequence, and vice versa. Following the assessment of fitness, the population was sorted in descending order of fitness to estimate the optimal answer. The optimal answer is the one that has the highest fitness. The Harris Hawks search the search space to estimate the ideal solution (prey) based on the identified best solution during the following phase of the HHO algorithm, called exploration. In hyperparameter optimization, the HHO method iteratively modifies each parameter's values by the best value discovered thus far in order to explore the search range of each parameter and determine its optimal value. The algorithm chooses a mild besiege step if the best solution found is good but could use some improvement. The stability and efficacy of the model are maintained as the parameter sequences are gradually updated in this step. Consequently, the algorithm chooses a hard besiege strategy if the predicted best solution is not good. To find the ideal value for the parameter solution, this technique updated it closely. The two other steps soft besiege with Rapid Dives and hard besiege with Rapid Dives do not apply to the work that is being proposed since they aggressively refine the solution of the parameters,

which could lower the stability and reliability of the model in practical situations. The mathematical definitions of the parameter solution updating employing the soft besiege and harsh besiege steps are found in Eqs. (7) and (8).

$$hy(t + 1) = \Delta hy(t) - F' |hy'(t) - hy(t)| \quad (7)$$

$$hy(t + 1) = hyt - F' \Delta hyt \quad (8)$$

where  $hy(t)$  denotes the current solution,  $hy(t + 1)$  represents the updated solution, and  $\Delta hy(t)$  indicates the deviation between the current and determined best solution. The fitness was assessed following the update of the solution to see if the resultant solution was ideal. During RBFNN training, the updated solution was applied if the computed new fitness was higher than the old fitness, and vice versa. This ongoing optimization of RBFNN hyperparameters improves overall detection accuracy as well as model training.

This hyperparameter optimization continues until reaching the maximum iteration count, and at each iteration, the HHO algorithm returns the best solution. Thus, the combination of RBFNN and HHO algorithm offers accurate and reliable detection of Alzheimer's disease using brain MRI images.

## 5 Results and Discussion

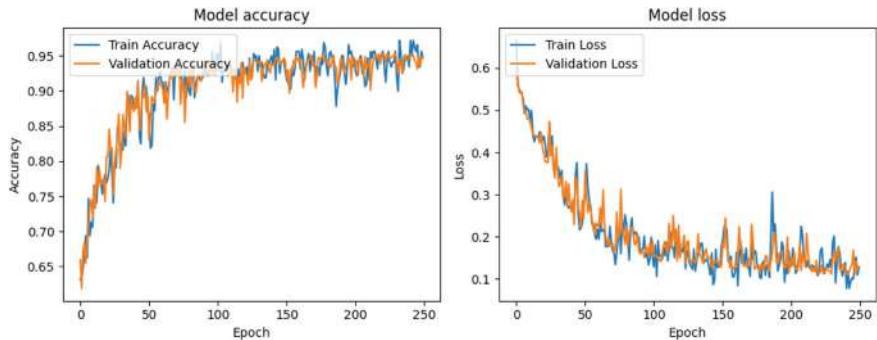
A hybrid Alzheimer's disease detection strategy was proposed in this study by combining the benefits of RBFNN and the HHO algorithm. The presented study was validated across ADNI and OASIS databases, and the results were determined in terms of accuracy, precision, recall, and f-measure.

### 5.1 Training and Testing Analysis

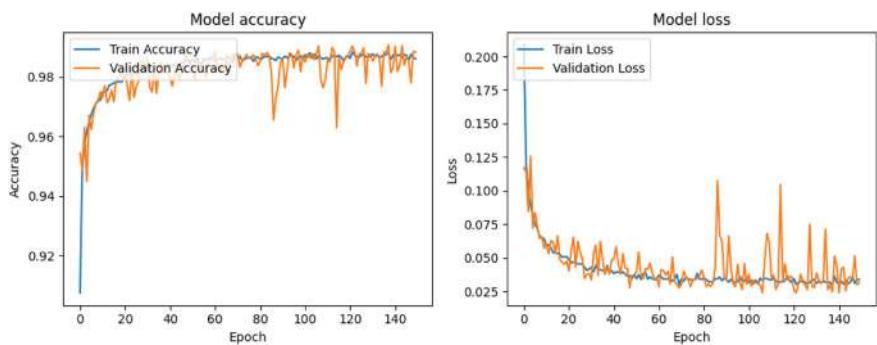
The training and testing performances of the suggested model throughout a series of increasing epochs are assessed in this subsection. First, the training and testing ratios of the input databases such as ADNI and OASIS were split into 70:30 for the model, and the results were assessed using metrics like accuracy and loss.

The testing accuracy quantifies how well the model applies to new brain MRI images, which are unseen data samples. The model's testing and training results for the ADNI dataset are shown in Fig. 4. The created algorithm successfully learns the patterns of normal and AD cases and generalizes on new images, as evidenced by the improvement in training and testing accuracy across increasing epochs.

As a result, during the training and testing stages, the loss measure was additionally assessed. The difference between the expected and actual outcomes on the known



**Fig. 4** Training and testing performance of ADNI dataset



**Fig. 5** Training and testing performance for OASIS database

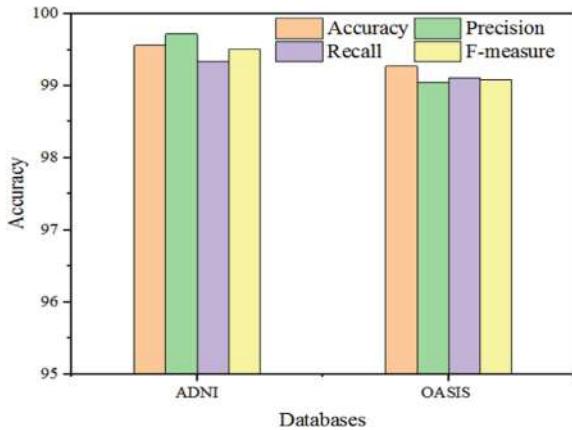
data is calculated using the training loss. On the other hand, the testing loss measures the variation between the actual and anticipated results on unidentified data samples. Moreover, the decrease in training and testing for increasing epochs indicates that the proposed technique successfully identifies AD from normal occurrences. In addition, the proposed technique effectively prevents overfitting, increasing its dependability for AD detection. Figure 5 presents the model testing and training outcomes for the OASIS database.

## 5.2 Performance Analysis

The recall, f-measure, accuracy, and precision are the performance measures utilized for analyzing the effectiveness of AD. Here, the experimentation is done using two databases, such as OASIS and ADNI. Additionally, the RBFNN-HHO model is trained and tested using these two databases.

**Table 1** Performance of the proposed strategy for two datasets

Databases	Performance (%)			
	Accuracy	Precision	Recall	F-measure
ADNI	99.56	99.72	99.34	99.51
OASIS	99.27	99.05	99.11	99.08

**Fig. 6** Assessment of RBFNN-HHO using ADNI and OASIS

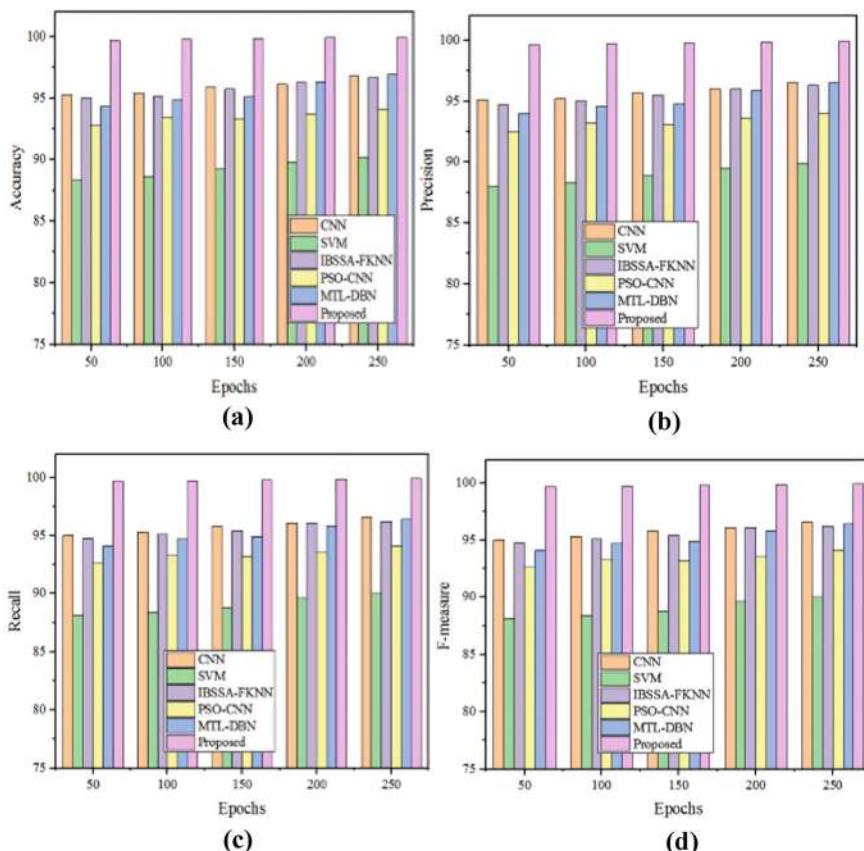
The RBFNN-HHO yielded 99.27% accuracy, 99.05% precision, 99.11% recall, and 99.08% f-measure using OASIS; whereas 99.56% accuracy, 99.72% precision, 99.34% recall, and 99.51% f-measure achieved by RBFNN-HHO using the ADNI database. Table 1 lists the effectiveness of the RBFNN-HHO using OASIS and ADNI datasets. Figure 6 shows the comparative performance of the RBFNN-HHO using the datasets. From the result, the ADNI database produced better performance for every metrics when compared to the OASIS dataset. This shows that the ADNI database contains significant qualities that enhance the learning and generalization capacities of the RBFNN-HHO, which leads to a more reliable and accurate diagnosis of AD.

### 5.3 Comparative Assessment

This section shows the effectiveness of the RBFNN-HHO by comparing its performance to the existing methods for AD detection. Some of the traditional models used for comparative evaluation are convolutional neural networks (CNNs) [30], support vector machines (SVM) [31], fuzzy k-nearest neighbor based on the improved binary salp swarm algorithm (IBSSA-FKNN) [32], particle swarm optimization with CNN (PSO-CNN) [33], and multi-task learning-based deep belief networks (MTL-DBN) [34]. The outcomes of the existing models validated using the Python language with ADNI and OASIS databases.

The RBFNN-HHO overall performance in identifying AD from brain MRI images is evaluated by the accuracy score. Here, two datasets covering increasing epochs are used to assess the accuracy of the established methods and the current model. The comparative performance of the suggested RBFNN-HHO algorithm for the ADNI dataset is shown in Fig. 7a-d. While the suggested technique produced an increased accuracy of 99.56% at 250th epoch, the current models including CNN, SVM, IBSSA-FKNN, PSO-CNN, and MTL-DBN obtained accuracy rates of 95.45%, 88.62%, 95.14%, 93.45%, and 94.85%, respectively.

Consequently, the accuracy was determined for different models for the OASIS database. At the 250th epoch, the existing and the proposed algorithms achieved an accuracy of 91.11%, 87.32%, 94.14%, 92.77%, 93.12%, and 99.27%. The significant improvement of accuracy by the developed algorithm highlights its effectiveness and robustness in Alzheimer's disease prediction compared to the existing models.



**Fig. 7** Comparison of model's performance for ADNI database: **a** Accuracy, **b** precision, **c** recall, and **d** f-measure

The robustness of the model in correctly identifying true positive instances is measured by the precision metric, which was also used to compare the precision performance of the proposed strategy with the existing models. For the ADNI dataset, the proposed approach and the conventional models mentioned above earned precision values of 95.91, 88.22%, 95.72%, 93.32%, 94.11%, and 99.56%, respectively, at the 250th epoch; however, for the OASIS database, these models earned precision values of 91.75%, 87.88%, 94.92%, 93.21%, 93.71%, and 99.05%, respectively.

When compared to traditional models, the suggested methodology's increased precision shows that it correctly detects actual positive instances. Improving the model's accuracy over time also shows how scalable and dependable it is in real-world situations.

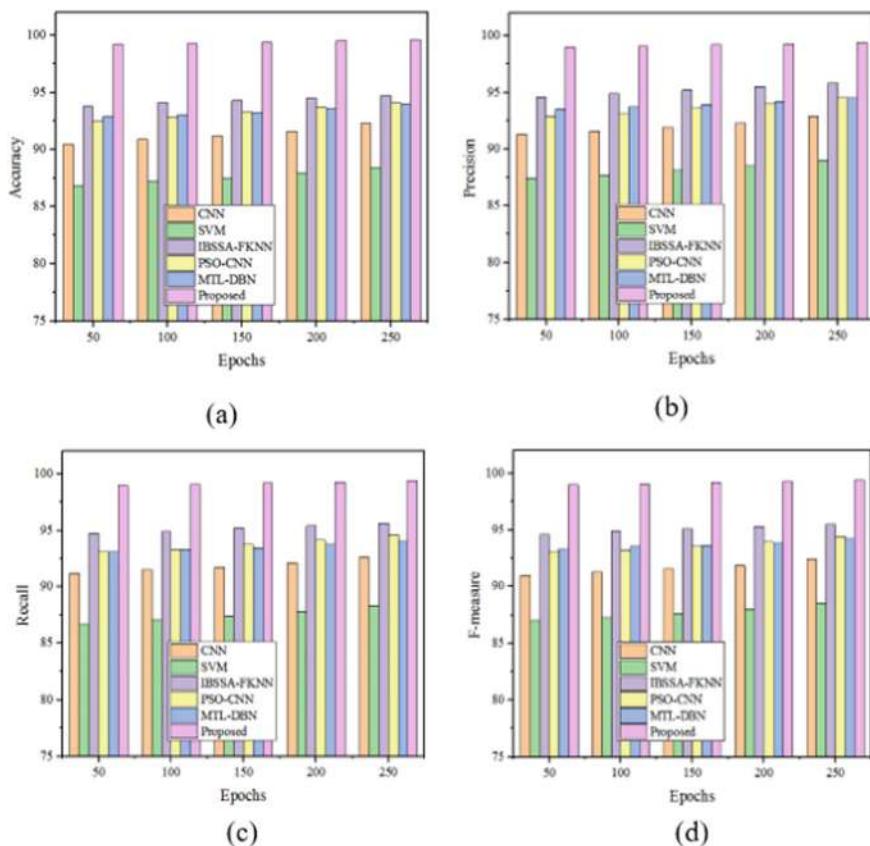
As a result, a comparison and validation of the recall performance were conducted using the current models. The recall metric measures how well the model finds all pertinent characteristics required for Alzheimer's disease prediction. The aforementioned conventional models produced recall rates for the ADNI database at the 250th epoch that were 96.12%, 88.80%, 96.11%, 93.75%, 94.35%, and 99.34%, respectively. Conversely, using the OASIS dataset at the 250th epoch, they obtained recall values of 91.45, 87.17, 95.17, 93.45, 93.37, and 99.11%. The suggested method's robustness in correctly diagnosing Alzheimer's disease is demonstrated by the considerable recall boost it experienced for the ADNI and OASIS datasets. The performance of RBFNN-HHO is compared with the current models on the OASIS dataset in Fig. 8a-d.

Concurrently, the suggested approach's f-measure performance was verified and contrasted with the current models. The average f-measures obtained by the available algorithms, namely CNN, SVM, IBSSA-FKNN, PSO-CNN, and MTL-DBN, for the ADNI dataset were 96.04%, 88.21%, 95.98%, 93.45%, 94.24%, and 99.51%, in that order. For the OASIS dataset, by contrast, these models yielded an average f-measure of 91.24%, 87.48%, 95.04%, 93.60%, 93.53%, and 99.08, respectively. But for the ADNI and OASIS databases, the developed method produced a higher average f-measure.

The suggested technique appears to provide a balanced performance in identifying both AD and normal cases based on the improvement of the f-measure. The average performance of various models for the ADNI dataset is compared in Table 2. The comparative evaluation of the average performance of the suggested approach and the current algorithm for the OASIS dataset is shown in Table 3.

It is evident from the thorough evaluation that, on a variety of datasets, including ADNI and OASIS, the created RBFNN-HHO algorithm performed better than the traditional models in terms of accuracy, precision, recall, and f-measure. The RBFNN-HHO algorithm is a highly useful tool for clinical applications because of its robustness, scalability, and dependability in effectively detecting Alzheimer's disease, as demonstrated by these data.

The proposed study hasn't included the clinical data of the patients in the dataset. By integrating the clinical dataset along with MRI will be an enhancement to the proposed methodology.



**Fig. 8** Comparison of model's performance for OASIS database: **a** Accuracy, **b** precision, **c** recall, and **d** f-measure

**Table 2** Comparative analysis of the model's performance for the ADNI dataset

Metrics	Average performances (%)					
	CNN	SVM	IBSSA-FKNN	PSO-CNN	MTL-DBN	Proposed
Accuracy	91.11	87.32	94.14	92.77	93.12	99.27
Precision	91.75	87.88	94.92	93.21	93.71	99.05
Recall	91.45	87.17	95.17	93.45	93.37	99.11
F-measure	91.24	87.48	95.04	93.60	93.53	99.08

**Table 3** Comparative assessment of the model's performance for the OASIS dataset

Metrics	Average performances (%)					
	CNN	SVM	IBSSA-FKNN	PSO-CNN	MTL-DBN	Proposed
Accuracy	91.11	87.32	94.14	92.77	93.12	99.27
Precision	91.75	87.88	94.92	93.21	93.71	99.05
Recall	91.45	87.17	95.17	93.45	93.37	99.11
F-measure	91.24	87.48	95.04	93.60	93.53	99.08

## 6 Conclusion

This paper presents the RBFNN-HHO for AD detection based on brain MRI images. The RBFNN-HHO enhances the accuracy and reliability of AD detection by integrating the benefits and effectiveness of RBFNN and HHO. Moreover, the images are pre-processed, and then data splitting is done. Then, the detection of AD is carried out using optimization and RBFNN. The results shows that the RBFNN-HHO performed better on both datasets. For the ADNI dataset, the RBFNN-HHO produced results of 99.56% accuracy, 99.72% precision, 99.34% recall, and 99.51% f-measure. Similarly, 99.27% accuracy, 99.05% precision, 99.11% recall, and 99.08% f-measure were obtained with the OASIS dataset. Furthermore, the RBFNN-HHO outperformed the state-of-the-art techniques, such as CNN, SVM, IBSSA-FKNN, PSO-CNN, and MTL-DBN on both datasets in terms of accuracy, precision, recall, and f-measure. These improved results show the recommended algorithm effectiveness and reliability for clinical applications and accurate diagnosis of Alzheimer's disease.

## References

1. Jiwtoode U, Chakole S, Bhatt N (2021) Alzheimer's disease: history, stages, diagnosis and its future. *J Pharm Res Int* 33(39A):41–45
2. McKeague B, Maguire R (2021) “The effects of cancer on a family are way beyond the person who's had it”: The experience and effect of a familial cancer diagnosis on the health behaviours of family members. *Eur J Oncol Nurs* 51:101905
3. Sharma S, Guleria K, Tiwari S, Kumar S (2022) A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer disease using MRI scans. *Measurement: Sens* 24:100506
4. Wang K, Tepper JE (2021) Radiation therapy-associated toxicity: Etiology, management, and prevention. *CA: A Cancer J Clin* 71(5), 437–454
5. Rossini PM, Di Iorio R, Vecchio F, Anfossi M, Babiloni C, Bozzali M, Bruni AC, Cappa SF, Escudero J, Fraga FJ, Giannakopoulos P (2020) Early diagnosis of Alzheimer's Disease: the role of biomarkers including advanced EEG signal analysis. Report from the IFCN-Sponsored Panel of Experts. *Clin Neurophys* 131(6):1287–1310
6. Tanveer M, Goel T, Sharma R, Malik AK, Beheshti I, Del Ser J, Suganthan PN, Lin CT (2024) Ensemble deep learning for Alzheimer's disease characterization and estimation. *Nat Mental Health*, pp 1–13

7. Hendi AA (2023) Development of medical imaging techniques: a review of technological advances in the field of medical imaging, such as magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound imaging (Ultrasound). *J Posit Psychol Wellbeing* 7(3):569–578
8. Mehmood A, Yang S, Feng Z, Wang M, Smadi Ahmad AL, Khan R, Maqsood M, Yaqub M (2021) A transfer learning approach for early diagnosis of Alzheimer's disease on MRI images. *Neuroscience* 460:43–52
9. Maryada SKR (2023) Application of deep learning to optimize computer-aided-detection and diagnosis of medical images (2023)
10. El-Sappagh S, Saleh H, Ali F, Amer E, Abuhmed T (2022) Two-stage deep learning model for Alzheimer's disease detection and prediction of the mild cognitive impairment time. *Neural Comput Appl* 34(17):14487–14509
11. Chang Z, Zhen D, Zhang F, Huang F, Chen J, Li W, Guo Z (2020) Landslide susceptibility prediction based on remote sensing images and GIS: Comparisons of supervised and unsupervised machine learning models. *Remote Sensing* 12(3):502
12. Yakimovich A, Beaugnon A, Huang Y, Ozkirimli E (2021) Labels in a haystack: approaches beyond supervised learning in biomedical applications. *Patterns* 2(12)
13. Ebrahimighahnavieh MA, Luo S, Chiong R (2020) Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review.“ Computer methods and programs in biomedicine 187 (2020):105242
14. Al-Shoukry S, Rassem TH, Makbol NM (2020) Alzheimer's diseases detection by using deep learning algorithms: a mini-review. *IEEE Access* 8:77131–77141
15. Usha V, Rajalakshmi NR (2023) Insights into diabetes prediction: a multi-algorithm machine learning analysis. In: 2023 4th international conference on smart electronics and communication (ICOSEC). IEEE, pp 1207–1212
16. Amiri Z, Heidari A, Navimipour NJ, Unal M, Mousavi A (2024) Adventures in data analysis: A systematic review of Deep Learning techniques for pattern recognition in cyber-physical-social systems. *Multimedia Tools Appl* 83(8):22909–22973
17. Gheisari M, Ebrahimzadeh F, Rahimi M, Moazzamigodarzi M, Liu Y, Pramanik PKD, Heravi MA et al (2023) Deep learning: applications, architectures, models, tools, and frameworks: a comprehensive survey. *CAAI Trans Intell Technol* 8(3):581–606
18. Cabitz F, Campagner A, Soares F, García de Guadiana-Romualdo L, Challa F, Sulejmani A, Seghezzi M, Anna Carobene. In: The importance of being external. Methodological insights for the external validation of machine learning models in medicine. *Comput Methods Prog Biomed* 208(2021):106288
19. Kieu ST, Hwa AB, Hijazi MHA, Kolivand H (2020) A survey of deep learning for lung disease detection on medical images: state-of-the-art, taxonomy, issues and future directions. *J Imag* 6(12):131
20. Castiglioni I, Rundo L, Codari M, Di Leo G, Salvatore C, Interlenghi M, Gallivanone F, Cozzi A, Claudia D'Amico N, Sardanelli F (2021) AI applications to medical images: From machine learning to deep learning. *Phys Medica* 83:9–24
21. Janghel RR, Rathore YK (2021) Deep convolution neural network based system for early diagnosis of Alzheimer's disease. *Irbm* 42(4):258–267
22. Sathiyamoorthi V, Ilavarasi AK, Murugeswari K Ahmed ST, Aruna Devi B, Kalipindi M (2021) A deep convolutional neural network based computer aided diagnosis system for the prediction of Alzheimer's disease in MRI images. *Measurement* 171:108838
23. Nguyen M, He T, An L, Alexander DC, Feng J, Thomas Yeo BT, Alzheimer's disease neuroimaging Initiative. Predicting Alzheimer's disease progression using deep recurrent neural networks. *NeuroImage* 222:117203
24. Dua M, Makhija D, Manasa PYL, Mishra P (2020) A CNN–RNN–LSTM based amalgamation for Alzheimer's disease detection. *J Med Biol Eng* 40(5):688–706
25. Umme Habiba S et al (2023) Transfer learning-assisted DementiaNet: a four layer deep CNN for accurate Alzheimer's disease detection from MRI images. In: Liu F, Zhang Y, Kuai H, Stephen EP, Wang H (eds) Brain informatics. BI 2023. Lecture Notes in Computer Science, vol 13974. Springer, Cham

26. The Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset accessible at [https://adni.loni.usc.edu/data-samples/access-data/#access\\_data](https://adni.loni.usc.edu/data-samples/access-data/#access_data)
27. The Open Access Series of Imaging Studies (OASIS) is accessible at <https://www.kaggle.com/datasets/ninadaithal/imagesoasis>
28. Wu J, Fang L, Dong G, Lin M (2023) State of health estimation of lithium-ion battery with improved radial basis function neural network. Energy 262:125380
29. Al-Betar MA, Awadallah MA, Makhadmeh SN, Doush IA, Abu Zitar R, Alshathri S, Elaziz MA (2023) A hybrid Harris Hawks optimizer for economic load dispatch problems. Alexandria Eng J 64:365–389
30. AlSaeed D, Omar SF (2022) Brain MRI analysis for Alzheimer's disease diagnosis using CNN-based feature extraction and machine learning. Sensors 22(8):2911
31. Raees, PCM, Thomas V (2021) Automated detection of Alzheimer's disease using deep learning in MRI. J Phys: Conf Ser 1921(1):012024
32. Lu D, Yue Y, Zhongyi H, Minghai X, Tong Y, Ma H (2023) Effective detection of Alzheimer's disease by optimizing fuzzy K-nearest neighbors based on salp swarm algorithm. Comput Biol Med 159:106930
33. Ibrahim R, Ghnemat R, Abu Al-Haija Q (2023) Improving Alzheimer's disease and brain tumor detection using deep learning with particle swarm optimization. AI 4(3):551–573
34. Zeng N, Li H, Peng Y (2023) A new deep belief network-based multi-task learning for diagnosis of Alzheimer's disease. Neural Comput Appl 35(16):11599–11610

# Assessing Wi-Fi Fingerprinting for Improved Indoor Positioning in Campus Settings: A Swedish University Example



Rasmus Andersson, William Tagesson, and Rashid Ali

**Abstract** Wi-Fi fingerprinting indoor positioning systems (FP-IPSSs) using RSSI are essential for indoor location-based services where GPS fails. This study evaluates four KNN algorithms (Traditional, Regions-based, Weighted Average, and Median Filtering) for RSSI-based Wi-Fi FP-IPS at University West's campus. Metrics assessed include accuracy, precision, and computational cost. The Regions-based algorithm excelled with an average error of 5.2 meters and a prediction time of 0.01 seconds. In contrast, the Traditional algorithm had higher errors (18.4 meters) but similar efficiency (0.01 seconds). Weighted Average and Median Filtering algorithms offered a balance between accuracy and cost. These findings highlight the regions-based algorithm's efficiency and accuracy for real-world applications.

**Keywords** Wi-Fi fingerprinting · Indoor positioning system · Indoor localization · Wi-Fi

## 1 Introduction

GPS, offering accuracy within 5 meters, is impaired indoors by Non-Line-of-Sight barriers like buildings. To address this, indoor positioning systems (IPSSs) using technologies such as magnetic, infrared, ultrasonic, ultrawide band, Bluetooth, and Wi-Fi have been developed, with Wi-Fi being popular due to its existing infrastructure [1].

---

This work was part of a project within the Wireless @ West Networking Research Group at the Department of Engineering Science, University West, Sweden. We thank OpenAI's ChatGPT for improving our writing clarity and accuracy.

R. Andersson · W. Tagesson · R. Ali (✉)

Department of Engineering Science, University West, Trollhättan, Sweden

e-mail: [rashid.ali@hv.se](mailto:rashid.ali@hv.se)

R. Andersson

e-mail: [rasmus.andersson.5@student.hv.se](mailto:rasmus.andersson.5@student.hv.se)

W. Tagesson

e-mail: [william.tagesson@student.hv.se](mailto:william.tagesson@student.hv.se)

In Wi-Fi-based IPS, the Received Signal Strength Indicator (RSSI) fingerprinting method is commonly used. This approach involves capturing signal data from access points (APs) and comparing it against an RSSI-based fingerprint database to estimate the user's position [2].

Various algorithms can be used to develop an RSSI-based Wi-Fi IPS, with a notable example being the K-nearest neighbor (KNN) algorithm. KNN operates by selecting the K-nearest reference points (RPs) and calculating the Euclidean distance to determine the closest RP [3]. KNN-based algorithms are popular due to their simplicity and scalability, making them easy to implement with existing RP datasets through numerical analysis. This study focuses on evaluating four distinct variations of KNN algorithms to assess IPS performance metrics, including accuracy, precision, and computational cost in a campus environment. The primary research questions are:

- How do different KNN algorithm variations perform on campus?
- Which Wi-Fi fingerprinting algorithm shows optimal performance in this setting?

## 2 Related Research

Wi-Fi fingerprinting uses captured RSSI values to locate users by comparing data within a fingerprint database. Basri et al. [4] outline IPS construction, evaluating Wi-Fi and Bluetooth technologies and emphasizing Wi-Fi fingerprinting's essential stages. Hu et al. [5] introduce Self-Adjusted Weight KNN (SAWKNN), demonstrating its superior performance. Hoang et al. [6] propose Soft Range Limited KNN (SRL-KNN), improving accuracy with user movement constraints. Lee et al. [7] utilize a random forest algorithm, employing smart watch data for indoor localization. Turabieh and Sheta [8] introduce Layered-Recurrent Neural Network (L-RNN), achieving superior accuracy with Nonlinear Regression (NLR). Quezada-Gaibor et al. [9] propose a data cleansing algorithm for Wi-Fi fingerprinting datasets, reducing positioning errors efficiently. These studies contribute to advancing Wi-Fi fingerprinting and indoor localization, enhancing accuracy and efficiency in real-world environments.

## 3 System Design

### 3.1 Campus Environment

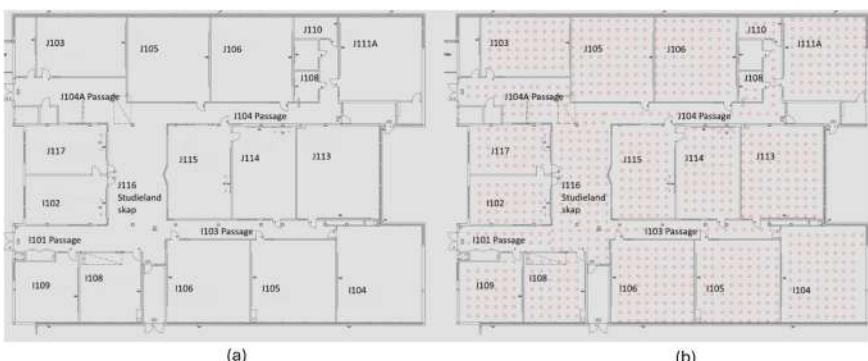
The study collected fingerprinting data within the City Campus of University West (Högskolan Väst, HV), Sweden, focusing on Level 1 of Block I and Block J. This area represents a typical indoor setting on the campus, enabling systematic Wi-Fi

signal data gathering to construct the fingerprint database. The campus map (Fig. 1a) highlights the specific locations where fingerprinting RPs were recorded (shown as red points in Fig. 1b), providing a foundational dataset for subsequent IPS analysis and experimentation.

### 3.2 Fingerprint Dataset

The initial phase involves gathering fingerprint data to establish a comprehensive database. Here, a fingerprint refers to the distinctive set of characteristics or attributes representing RSSI values or sensors at precise locations within a building. Multiple RPs constitute the fingerprint database, utilized by algorithms to determine a device's location. Essential data captured include RSSI and the unique identifier (BSSID). RSSI-based Wi-Fi fingerprinting typically comprises offline and online phases.

**Offline Phase:** During the offline phase, fingerprints were systematically collected to construct the database. This involves collecting RPs in classrooms and corridors, spaced one meter apart, to build a fingerprinting database. A Galaxy A52 5G smartphone with the GetSensorData 2.0 application facilitated data recording [10]. Data collection occurred for about 20s at each RP, using a systematic sampling approach. RPs were gathered on the first floor of Blocks I and J of the HV campus, as illustrated in Fig. 1a, with corresponding real-world positions recorded. The dataset and maps are available on our GitHub repository [11]. After preprocessing the raw data, fingerprints from each RP were used to build the database. RPs, representing real-world positions, were mapped onto pixels corresponding to the layout of the first floor of Blocks I and J of the HV campus. The map dimensions are  $1920 \times 1356$  pixels, providing approximately 35.7 pixels per meter for precise spatial alignment



**Fig. 1** Campus environment and fingerprinting reference points. **a** Map depicting the campus environment encompassing Level 1 of Block I and Block J at the City Campus of University West, Sweden. **b** The same map highlighting fingerprinting Reference Points (RPs) collected within this environment, denoted by red points

(Fig. 1a). With this conversion, RPs can accurately be depicted on the map, as shown in Fig. 1b.

**Online Phase:** In the online fingerprinting phase, test data points are collected for testing purposes, compared against the fingerprint database. Four distinct KNN algorithms (refer to Sect. 3.3) are used for positional estimation of these test points (TPs). We collected TPs using the same application used for RP collection. Unlike the systematic sampling for RPs, TP data are collected at random locations, ensuring that TPs do not overlap with RPs.

### 3.3 Indoor Positioning Algorithms (*K*-Nearest Neighbor)

The KNN algorithm computes device proximity to predefined points by selecting the  $K$ -nearest neighbors and calculating distances. In Wi-Fi fingerprint IPS, it identifies the nearest RPs to TPs, providing localization. Four KNN variants were evaluated: Traditional, Weighted Average, Median Filtering, and Regions-based, the latter offering an enhanced solution for campus environments.

**Traditional KNN Algorithm** Modern KNN IPS algorithms employ diverse learning methods to predict the optimal RP, requiring a sizable and homogeneous dataset. Traditional KNN algorithms rely on fixed calculations, resulting in consistent outcomes without learning. They take online data for a specific TP and the entire fingerprint database as inputs, iterating through RP positions and BSSIDs. By comparing BSSIDs and calculating RSSI differences, they determine multiple candidate positions based on RSSI errors, selecting the candidate with the lowest overall error.

**Weighted Average KNN Algorithm** The Weighted Average variant prioritizes earlier BSSID findings based on  $K$ , utilizing descending RSSI-ordered datasets. It employs a weighted approach inversely proportional to error, preventing division by zero and assigning higher weights to smaller errors. The algorithm calculates the sum of weighted error RSSI values and divides by the sum of reciprocals of the weights, utilizing a “data overlapping” technique.

**Median Filtering KNN Algorithm** This method employs a sliding window algorithm for filtering and optimizing outcomes. Consecutive BSSIDs overlap, enhancing position estimation and error averaging. The window comprises five loops, with every fifth loop incorporating overlapping data from the previous loop, while others add current index data. Utilizing the data overlapping generates substantial data. In median filtering, RSSI values are averaged using a window of RSSI values, from which the median is calculated (filtered values). Each error value within the window is compared to its corresponding filtered value to ensure proximity to measured actual values. The algorithm retains actual values and filters incorrect ones.

### 3.3.1 Regions-Based KNN Algorithm

Partitioning the experimental area into distinct regions enhances accuracy by limiting the number of accessible RPs to a TP. Previously, TPs could select from all RPs, but with this modification, computational time is expected to decrease and large-scale errors should diminish. Regions are delineated based on the BSSID broadcasted from surrounding APs (in our case, six) of all RPs. Each RP is categorized into a region based on the BSSID with the highest RSSI value, assuming that the corresponding AP belongs to that region. Subsequently, all RPs are categorized into one of the six regions, with data separated accordingly. Before predicting the RP to which a TP belongs, the TP is assigned a region using the same method as the RPs. The algorithm then executes the default KNN algorithm, as mentioned in Sect. 3.3, with the TP selecting from RPs within the same region.

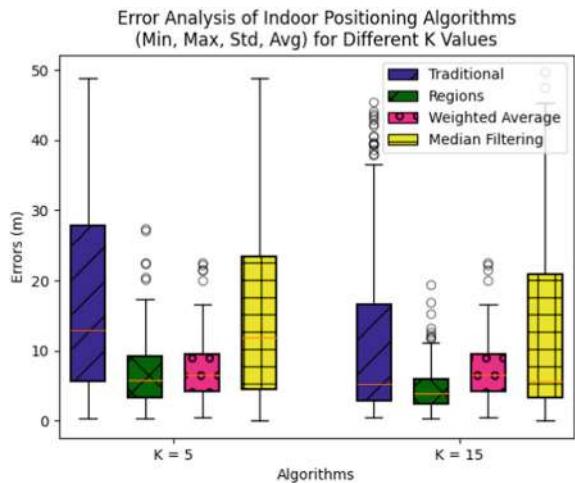
## 4 Performance Evaluation

For performance evaluation, accuracy, precision, and computational cost were selected as metrics. Accuracy measures the proximity of positioning results to the actual location of a TP, averaged across error measurements for each TP. Precision evaluates the consistency of distinct variations of KNN algorithms across all tests, regardless of individual test accuracy. Cumulative Distribution Function (CDF) and quartile values visualize error probability and distribution. Computational cost measures the time needed to compute results for one TP, represented as the average time from start to finish for all TPs in milliseconds, varying based on the four KNN algorithms and  $k$  value.

In Fig. 2, the box plot illustrates the error distribution in meters for the four KNN algorithms at two different values of  $k$  (5 and 15). For  $k = 5$ , the Traditional algorithm shows the highest error variability and median error. The Regions-based algorithm demonstrates lower errors and variability compared to the Traditional method. The Weighted Average algorithm exhibits slightly higher errors than the Regions-based method but lower than the Traditional approach. The Median Filtering algorithm shows moderate accuracy but with a wider spread. For  $k = 15$ , the Traditional algorithm's errors decrease compared to  $k = 5$  but still show considerable variability. The Regions-based algorithm maintains the lowest error distance among the four methods. The Weighted Average algorithm has similar performance to the Regions-based method but with slightly higher errors. The Median Filtering algorithm displays increased errors and variability compared to the other methods. Outliers are present in all algorithms, indicating occasional significant deviations from the typical error range. The Regions-based algorithm consistently performs best in terms of lower average errors and variance.

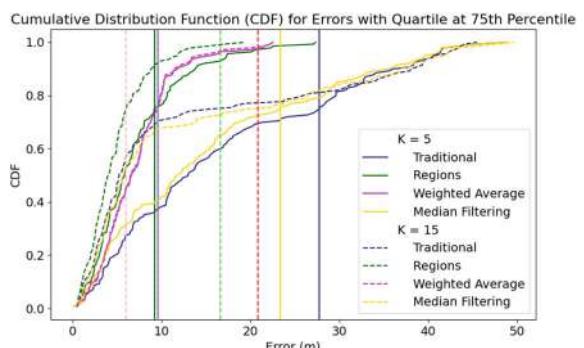
The CDF (Fig. 3) plot shows the cumulative probability distribution of accuracy errors to demonstrate the precision of four KNN algorithms at  $k = 5$  and  $k = 15$ . For  $k = 5$ , the Traditional algorithm (solid blue line) exhibits the slowest increase

**Fig. 2** Box plot showing the accuracy errors (m) of four KNN algorithms (Traditional, Regions-based, Weighted Average, and Median Filtering) for  $k = 5$  and  $k = 15$ . The plot includes average and variance values with outliers



in the CDF, indicating higher errors. The Regions-based algorithm (solid green line) shows the steepest curve, indicating lower errors and better performance. The Weighted Average algorithm (solid pink line) also performs well, closely following the Regions-based algorithm. The Median Filtering algorithm (solid yellow line) has a moderate performance, with errors higher than the Regions-based and Weighted Average but lower than the Traditional algorithm. For  $k = 15$ , the Traditional algorithm (dashed blue line) shows improved performance compared to  $k = 5$  but still has higher errors than the other methods. The Regions-based algorithm (dashed green line) continues to show the best performance with the lowest errors. The Weighted Average algorithm (dashed pink line) maintains good performance, similar to the Regions-based method. The Median Filtering algorithm (dashed yellow line) shows slightly decreased errors compared to  $k = 5$ . The vertical dashed lines represent the 75<sup>th</sup> percentile (Q3) for each algorithm, indicating the error value below which 75% of the data points fall. The Regions-based algorithm consistently shows the lowest

**Fig. 3** Cumulative distribution function including 75th percentile (Q3) of accuracy errors for KNN algorithms with  $k = 5$  and  $k = 15$



Q3 values, highlighting its superior accuracy in reducing errors compared to the other methods.

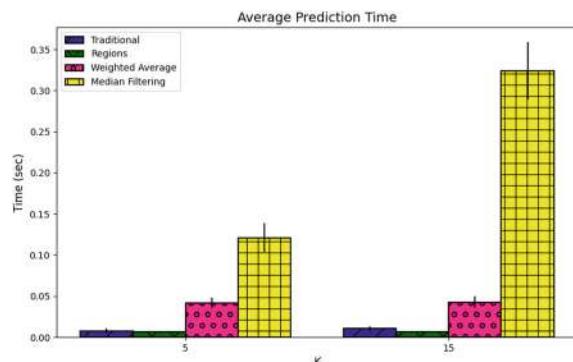
The bar chart in Fig. 4 shows the average prediction time (in seconds) for the algorithms. The Traditional algorithm has very low prediction times for both  $k = 5$  and  $k = 15$ , indicating high computational efficiency. However, the Regions-based algorithm demonstrates the lowest prediction times, lower than the Traditional algorithm. The Weighted Average algorithm has moderate prediction times, higher than the Traditional and Regions-based algorithms but still relatively efficient. The Median Filtering algorithm has the highest prediction times for both  $k = 5$  and  $k = 15$ , with a significant increase as  $k$  rises, indicating a higher computational cost. Overall, the Regions-based approach is as efficient as the Traditional method and outperforms the Weighted Average and Median Filtering methods in terms of prediction time, although the Median Filtering algorithm incurs a much higher computational cost. These results are crucial for drawing conclusions on algorithm performance for RSSI-based Wi-Fi fingerprinting IPSs.

Computational efficiency and prediction accuracy are vital considerations for real-world applications. The findings indicate that the Regions-based algorithm offers a compelling balance of low prediction time and high accuracy, making it a promising choice for practical implementation in indoor positioning systems in campus environments.

## 5 Conclusion

This study emphasizes the importance of tuning and evaluating Wi-Fi fingerprinting IPS algorithms, especially in campus environments where accurate localization is critical. We assessed various KNN algorithms within University West's indoor campus. Our findings show that the Regions-based algorithm consistently outperforms Traditional, Weighted Average, and Median Filtering algorithms in minimizing accuracy errors and prediction time. These results improve indoor positioning accuracy

**Fig. 4** Average prediction time of each algorithm with different  $k$  (5 and 15) values, highlighting the computational efficiency of the regions-based approach



and offer valuable insights for engineers implementing Wi-Fi FP-IPS, highlighting the regions-based algorithm's suitability for real-world applications.

## References

1. Qi L, Liu Y, Yu Y, Chen L, Chen R (2024) Current status and future trends of meter-level indoor positioning technology: a review. *Remote Sens* 16(2). <https://doi.org/10.3390/rs16020398>, <https://www.mdpi.com/2072-4292/16/2/398>
2. Álvarez Merino CS, Khatib EJ, Luo-Chen HQ, Muñoz AT, Moreno RB (2024) Evaluation and comparison of 5g, wifi, and fusion with incomplete maps for indoor localization. *IEEE Access* 12:51893–51903. <https://doi.org/10.1109/ACCESS.2024.3384625>
3. Srivastava T (2024) A complete guide to K-Nearest neighbors (Updated 2024)—analyticsvidhya.com. <https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/>, [Accessed 25 Apr 2024]
4. BASRI C, El Khadimi A (2016) Survey on indoor localization system and recent advances of wifi fingerprinting technique. In: 2016 5th international conference on multimedia computing and systems (ICMCS), pp 253–259. <https://doi.org/10.1109/ICMCS.2016.7905633>
5. Hu J, Liu D, Yan Z, Liu H (2019) Experimental analysis on weight  $K$ -nearest neighbor indoor fingerprint positioning. *IEEE Internet Things J* 6(1):891–897. <https://doi.org/10.1109/JIOT.2018.2864607>
6. Hoang MT, Zhu Y, Yuen B, Reese T, Dong X, Lu T, Westendorp R, Xie M (2018) A soft range limited k-nearest neighbors algorithm for indoor localization enhancement. *IEEE Sens J* 24:10208–10216. <https://doi.org/10.1109/JSEN.2018.2874453>
7. Lee S, Kim J, Moon N (2019) Random forest and wifi fingerprint-based indoor location recognition system using smart watch. *Human-centric Comput Inf Sci* 9(1) (2019). <https://doi.org/10.1186/s13673-019-0168-7>
8. Turabieh H, Sheta A (2019) Cascaded layered recurrent neural network for indoor localization in wireless sensor networks. In: 2019 2nd international conference on new trends in computing sciences (ICTCS), pp 1–6. <https://doi.org/10.1109/ICTCS.2019.8923086>
9. Quezada-Gaibor D, Klus L, Torres-Sospedra J, Lohan ES, Nurmi J, Granell C, Huerta J (2022) Data cleansing for indoor positioning wi-fi fingerprinting datasets. In: 2022 23rd IEEE international conference on mobile data management (MDM), pp 349–354. <https://doi.org/10.1109/MDM55031.2022.00079>
10. Jiménez AR, Seco F, Torres-Sospedra J (2019) Tools for smartphone multi-sensor data registration and gt mapping for positioning applications. In: 2019 International conference on indoor positioning and indoor navigation (IPIN), pp 1–8. <https://doi.org/10.1109/IPIN.2019.8911784>
11. Andersson R, Tagesson W, Ali R (2024) Dataset: evaluating Wi-Fi fingerprinting for enhanced indoor positioning in campus environments. <https://github.com/wirelessATwest/PEWFIPS-HV>

# Optimization Algorithm of Blockchain Smart Contracts for Digital Economy



Zhen Zang

**Abstract** In response to the low efficiency and high cost of smart contract execution in the current digital economy environment, this article conducts algorithm optimization research based on blockchain smart contracts. Using Convolutional Neural Network (CNN) method, key features are extracted from massive data, and their historical running records are sequentially modeled. The behavior rules of contracts are analyzed to identify potential vulnerabilities and abnormal behaviors. Adopting the reinforcement learning algorithm of Deep Q-Network (DQN), the optimal decision is made based on the characteristics and behavior of smart contracts. This article compares the operating costs of two algorithms under the same contract, compares the amount of gas consumed by the two methods under the same contract, and their impact on system operating costs. It also compares them with existing dynamic scheduling methods to achieve the goal of reducing operating costs. The research results indicate that the optimization method proposed in this article can effectively reduce network operating costs and provide support for the construction of the national digital economy.

**Keywords** Digital economy optimization algorithm · Blockchain smart contracts · Efficiency enhancement · Security improvement · Privacy protection

## 1 Introduction

Based on blockchain technology, studying the optimization method of smart contracts based on blockchain is of great significance for promoting the development of the digital economy [1–3]. This article proposes a new method based on deep learning for contract execution path prediction, contract optimization strategy generation, and contract code compression. Traditional scheduling strategies have problems such as static scheduling and poor scalability. The existing static scheduling methods

---

Z. Zang (✉)

School of Economics and Management, Chengdu Technological University, Chengdu, China  
e-mail: [zangzhen0325@163.com](mailto:zangzhen0325@163.com)

cannot dynamically adjust the execution sequence of contracts, and the centralization of their execution structure limits their scalability. In response to this issue, this article conducts research from three aspects: execution efficiency, security, and scalability. This study proposes a flexible adjustment method suitable for various operational environments and can make appropriate adjustments based on actual operational conditions, thereby better adapting to the development needs of the digital economy [4]. This article is based on the digital economy and studies the optimization algorithm of smart contracts based on blockchain.

By analyzing smart contract data, the article extracted key information, extracted structural information and attributes of the contract, monitored the operational status of the contract, and extracted dynamic characteristic information of the system. By detecting potential security defects in the contract, it also evaluated the performance of the contract, and used recurrent neural network (RNN) to model the historical operational data of the contract in sequence. This article intends to use the DQN reinforcement learning algorithm to identify common and abnormal behaviors in contracts, detect defects and risks in contracts, and optimize the characteristics and behaviors of contracts using the DQN reinforcement learning algorithm. On this basis, reasonable state expression methods, action spaces, and reward functions were designed to ensure the efficient and secure execution of the contract. Finally, this article evaluates and records the contract execution time, calculates the average execution time, and conducts statistical analysis on the execution time of multiple contracts; this article uses gas consumption as an indicator to evaluate the economic efficiency of contract execution, and compares the gas consumption of optimization algorithms and traditional algorithms. It combines historical data analysis and vulnerability detection algorithms to evaluate contract security, identify potential security vulnerabilities, monitor abnormal behavior in contracts, and ensure the safe operation of contracts.

## 2 Related Works

Previous studies have mainly used dynamic scheduling algorithms to improve the execution efficiency of blockchain smart contracts in the digital economy. The dynamic scheduling algorithm aims to dynamically adjust task allocation and resource utilization based on real-time contract execution and system resource utilization [5, 6]. Researchers such as Dolgui A have developed and tested a new model for smart contract design in the supply chain, which allows for updating operational status in the blockchain using state control variables, thereby providing automatic information feedback, interrupt detection, and contract execution control [7]. Gao Z and other researchers parse the smart contract code into a word stream containing code structure information, transforming code elements (such as statements, functions) into numerical vectors that should encode the syntax and semantics of the code. He compared the similarity between the vectors of the encoded code and known errors to identify potential issues in the research [8]. The dynamic scheduling algorithm focuses on real-time task allocation and resource scheduling, which increases the

computational cost and time complexity of the system. Although it has real-time and flexibility, it is greatly affected by complexity and lacks reliability.

To overcome the limitations of traditional methods, some researchers have begun to explore the optimization of blockchain smart contract algorithm technology in the field of digital economy, using methods such as deep learning and neural networks. These technologies automatically adjust the execution mode of smart contracts by learning and simulating the patterns and patterns during the contract execution process, improving the efficiency and performance of smart contract execution. Zhang Y and other researchers proposed a construction site information management system framework based on blockchain and smart contracts. Multiple independent smart contracts have been developed for different data types, integrating different data analysis algorithms such as deep learning into information management and ensuring that data is not tampered with [9]. Wang W and other researchers extracted binary features from the simplified operation code of smart contracts, and then used five machine learning algorithms and two sampling algorithms to construct the model. They then evaluated 49,502 real-world smart contracts running on Ethereum [10]. Although deep learning and neural network technologies have made progress in improving the efficiency and security of contracts, they cannot meet the needs of the digital economy due to their poor interpretability and limited resources [11]. Based on the limitations of existing methods, this article adopts dynamic scheduling algorithms to optimize the engine performance of contract execution.

### 3 Building a Smart Contract Optimization Framework

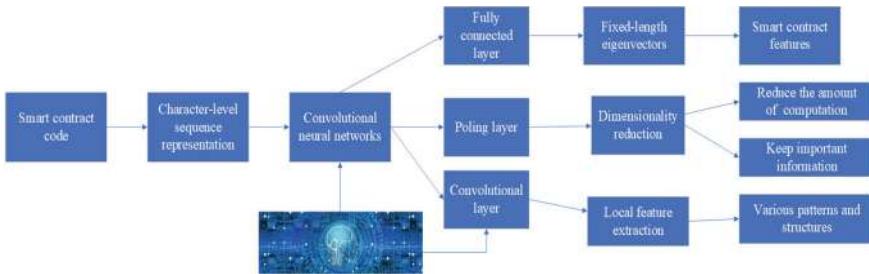
#### 3.1 *Smart Contract Feature Extraction*

Static analysis techniques can be used to extract the structural information and attributes of contracts, run contracts, and monitor the data and behavior generated during their execution, extracting feature information during the execution process. It detects potential security vulnerabilities in contracts, extracts security related features, evaluates contract execution time and gas consumption, and extracts performance related features.

Figure 1 illustrates the process of using CNN as a feature extractor to extract key features in smart contracts.

#### 3.2 *Analysis of Smart Contract Behavior*

Cyclic neural networks have been widely used in natural language processing, speech recognition and time series analysis. Smart contracts are contracts that are automatically enforced, with the buyer and seller writing the terms directly in code. Contracts



**Fig. 1** Feature extraction process diagram

are stored in the blockchain and are automatically enforced in certain situations. Using RNN to model the dynamic characteristics of contracts is a significant research topic. Each execution record contains information about the contract status at a given time, such as input arguments, output values, and all events raised during execution. On this basis, this paper presents a dataset-based approach to describe the contract changes in different periods.

A new recurrent neural network (GRU) model is proposed for efficient analysis and longtime dependence of time series data. A new recurrent neural network model is proposed to solve the problems in dealing with large time series. On this basis, join GRU, better capture the data association, thus improving the accuracy and efficiency of simulation. Once the recurrent neural network has been trained into a series of execution records, this method is used to perform intelligent contract operations and to detect anomalies. By contrasting the expected contract with the actual contract, anomalies are identified and possible weaknesses and dangers are identified. This helps contract developers and auditors identify problems early and avoid security breaches or financial losses.

RNN not only finds the abnormal phenomenon in the intelligent contract, but also predicts it according to the previous data. Through the study of the trends and patterns of business records, the performance of contracts under various circumstances can be reasonably predicted. This can be especially useful to optimize the execution of intelligent contracts and to increase the efficiency of the blockchain. This paper provides a basis for constructing smart contract mode with IPR and enhancing the safety and reliability of the contract. By analyzing the time sequence data efficiently and using genetic expression, it is possible to know more about the development process of smart contract and decrease its hidden risk. Recently, it has become very important for the safety and reliability of the distribution system to introduce NNN into the Intelligence Contract Analysis System.

### 3.3 Smart Contract Optimization Strategy Generation

The DQN method can be applied to produce optimal policies according to the features and actions of intelligent contracts. It can build up the Intelligent Contract Execution Environment Model, Set up Proper Status Representation, Movement Space and Reward Function, DQN is a Value Function Approximation to Continuous Interaction and Study of Best Policies. The optimal policy consists of optimal contract code, reasonable allocation of resources, etc. Furthermore, the optimal policy is used to improve the performance of smart contracts.

State representation function:

$$S = (f_1, f_2, \dots, f_n) \quad (1)$$

Among them,  $f_1$  represents the contract size, and  $f_n$  represents the nth characteristic.

Action space function:

$$A = \{a_1, a_2, \dots, a_m\} \quad (2)$$

Among them,  $a_m$  is the  $m$  th action.

Reward function:

$$R(s, a, s') = \alpha \cdot R_{\text{efficiency}}(s, a, s') + \beta \cdot R_{\text{security}}(s, a, s') - \gamma \cdot C(a) \quad (3)$$

Among them,  $R_{\text{efficiency}}(s, a, s')$  represents the security reward for smart contracts, while  $R_{\text{security}}(s, a, s')$  represents the security reward for only contracts. An important advantage of using reinforcement learning methods such as DQN to optimize intelligent contracts is that it has strong adaptability and reference. Classical optimization methods require prior knowledge of how the contract works or specific optimization criteria. On the other hand, reinforcement learning automatically adjusts the optimal decision through trial and error, which makes it more adaptive.

In order to make better use of the accumulative points network to optimize the contract, it is necessary to construct the operation environment model that can capture the contract elements. The model should include inputs, outputs, state transitions, and other external factors that affect the performance of the contract. On this basis, a new analysis method of contract behavior is proposed and the corresponding improvement method is given. Secondly, aiming at the optimal contract problem, the reasonable state expression and action space are studied. The state representative should capture information about the current state of the contract, such as its variable functions and execution. The action space should identify actions for optimal contracts, such as modifying contract codes, adjusting resource allocations, or changing contract parameters.

On this basis, this project proposes an improved learning method to optimize the contract efficiently. This reward function provides a feedback to the optimal decision quality and motivates the performance behavior to achieve the desired results. Through reasonable return function, the whole system can be guided to optimize, so as to reduce gas cost, maximize system efficiency, or enhance system safety.

Once the running environment of the system is constructed, the system is modeled, and then the revenue function is modeled to achieve optimal control of the system. DQN is a deep re-learning method, under certain conditions, the depth of the network and Q-learning organic integration, under certain conditions, automatically from the optimal decision. By modeling and rewarding the contract implementation environment of DQN agent, a multi-agent-based contract implementation scheme is formed to achieve efficient contract implementation. The optimal control method based on reinforcement learning, such as DQN, is a good solution. For example, the method realizes the discovery and repair of defects, the optimal allocation of resources and performance, and the streamlining of contract implementation process. DQN agent can adjust its optimal decision according to the change of external environment through continuous interaction with contract and self-cognition of contract behavior.

In terms of contract optimization, the reinforcement learning method can also be applied to the allocation and management of intelligent contracts in blockchain systems. For example, this algorithm helps determine the most cost-effective deployment contract, efficient allocation of resources, and overall contract performance. Block chain developers enhance the extensibility, security, and reliability of smart contracts through enhanced learning. In a word, applying the improved reinforcement learning technology (DQN) to the optimization of intelligent contracts is an effective way to improve the performance and operational efficiency of blockchain technology. This project will give full play to the self-adaptive and self-adaptive characteristics of the above methods, and design an optimal decision scheme in line with the characteristics of smart contract. Intelligent contract design based on reinforcement learning is an important way to improve the performance of distributed systems and enhance the overall user experience.

## 4 Results and Discussion

### 4.1 Evaluation of Smart Contract Execution Time

The average execution time of smart contracts can be used as an evaluation indicator to select from a batch of smart contracts, execute these contracts in the same testing environment, and record the execution time of each contract. The time module in Python can be used to accurately measure the execution time of contracts. After executing all contracts, the execution time of each contract can be statistically analyzed to calculate the average execution time. This article can conduct statistical analysis on the execution time of multiple contracts.

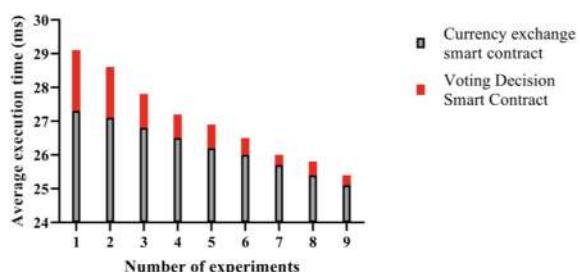
Table 1 shows the execution time and average execution time of voting system smart contracts, lottery smart contracts, decentralized exchange smart contracts, supply chain tracking smart contracts, insurance claims smart contracts, and crowd-funding smart contracts under the same number of executions. The average execution time of supply chain tracking smart contracts is the shortest, at 10 ms, which is the best in terms of execution efficiency.

Figure 2 shows the changes in the average execution time of currency exchange smart contracts and voting decision smart contracts. Multiple experiments can be conducted on these two contracts, and it can be observed that the average execution time decreases sequentially. The average execution time of currency exchange smart contracts in different experiments is 27.3, 27.1, 26.8, 26.5, 26.2, 26, 25.7, 25.4, and 25.1 ms. The voting decision smart contract is 29.1, 28.6, 27.8, 27.2, 26.9, 26.5, 26, 25.8, and 25.4 ms. A shorter average execution time means higher execution efficiency, indicating that currency exchange smart contracts and voting decision smart contracts have faster execution efficiency, and optimization algorithms perform better in improving the execution efficiency of smart contracts.

**Table 1** Smart contract execution data table

Smart contract	Execution time (ms)	Number of executions	Average execution time (ms)
Voting system smart contract	116	10	11.6
Lottery smart contract	250	10	25
Decentralized exchange smart contract	365	10	36.5
Supply chain tracking Smart contract	100	10	10
Insurance claim smart contract	448	10	44.8
Crowd funding smart contract	256	10	25.6

**Fig. 2** Execution time variation chart



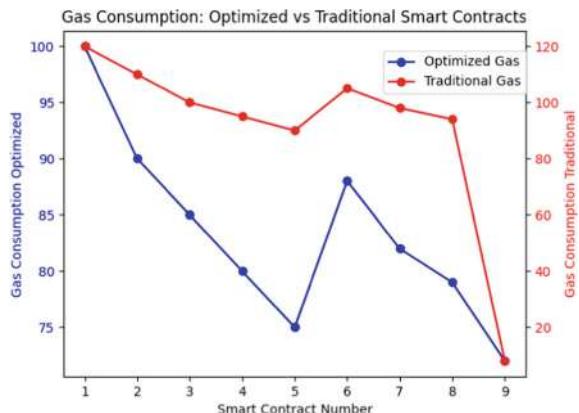
## 4.2 Execution Cost Assessment

Gas consumption can be used as an important indicator, and in blockchain systems, gas consumption reflects the economic cost of contract execution. To evaluate the effectiveness of the optimization algorithm and execute a batch of smart contracts, this article records their gas consumption in detail. It uses Solidity to write smart contracts, deploys and executes them in the Ethereum testing network, and captures the gas consumption during the contract execution process. It utilizes the Solidity integrated development environment to monitor the gas consumption of contracts in real time, ensuring the accuracy and reliability of data. By comparing the number of gas consumed by optimization algorithms and dynamic scheduling algorithms to execute the same contract, this article evaluates the impact of optimization algorithms on execution costs and comprehensively understands the effectiveness of optimization algorithms in improving the economic efficiency of contract execution.

Figure 3 shows a comparison of the number of gas consumed by optimization algorithms and dynamic scheduling algorithms in executing smart contracts. Through observation, it was found that the gas consumption of the optimization algorithm on different smart contract numbers is 100, 90, 85, 80, 75, 88, 82, 79, 72, all of which are lower than the dynamic scheduling algorithm's 120, 110, 100, 95, 90, 105, 98, 94, 85. The novel of gas consumption indicates that the economic cost of smart contracts is low, and the efficiency of contract execution is higher. The contract code can be more concise and efficient, reducing the economic cost of the contract. Through this comparison, the article has learned about the effectiveness of optimization algorithms in improving the economic efficiency of contract execution.

Revolutionizing business and smart contracts by providing a secure, transparent, decentralized platform. An intelligent contract is an automated contract in which the terms of the contract are written directly into the code. They automate the

**Fig. 3** Comparison of gas consumption



terms of contracts without intermediaries, thus increasing the effectiveness and cost-effectiveness of transactions. However, the effect and cost of smart contract implementation is an important issue related to its practicability and practicality. The purpose of this study is to compare the average execution time and energy consumption of smart contracts for cash exchange and voting decisions as shown in Figs. 2 and 3. On this basis, we will also explore the role of contract design method based on optimal policy in improving contract execution efficiency and reducing transaction costs.

Figure 2 shows the average elapsed time in several experiments where money was exchanged with a vote on a smart contract. The data show that the average performance cycle of the above two types of intelligent contracts shows a continuous downward trend. This trend shows that the optimal algorithm can improve the performance of intelligent contracts. In practical applications, because of the lower average speed and higher speed, it has a good application prospect. Taking foreign exchange trading as an example, simulation results show that the speed of the system is reduced from 27.3 ms to 25.1 ms. It is found that the execution time of intelligent contracts based on voting decision is reduced from 29.1 ms to 25.4 ms by comparing two different types of intelligent contracts.

Figure 3 shows the optimal and dynamic scheduling algorithms in the implementation of intelligent contracts air consumption rate comparison. In the blockchain system, gas consumption is a key factor to measure the cost of contract performance. The utility model reduces the gasoline consumption, reduces the economic cost and improves the implementation effect of the contract. The data shown in Fig. 3 show that the optimal performance method always dissipates more gas than a dynamic scheduling performance using various smart contract numbers. The air consumption calculated by the optimal method is 72–100, while the air consumption calculated by the dynamic planning method is 85–120. Experimental results show that this method can effectively reduce the cost of intelligent contract implementation. By optimizing the contract code, the execution speed of the contract is improved, and the intelligent contract has a better cost-performance ratio.

The optimal strategy proposed in this paper is of great significance to improve the effectiveness and economy of contract execution. On this basis, a multi-objective cooperative control method based on genetic network is proposed. In this way, in all walks of life, smart contracts become more practical and feasible. Therefore, the effective implementation of contract intelligence is the key to promote the application and development of blockchain technology. Through the use of optimal methods to improve work efficiency, reduce costs, so that intelligent contracts have higher effectiveness, cost-effectiveness, and scalability. On this basis, this project proposed a new network model based on network model, and used this model to establish the corresponding network model to improve the performance of the model.

## 5 Conclusions

This article focuses on the optimization of blockchain smart contracts in the field of digital economy. The use of blockchain smart contract optimization algorithms, combined with parallel computing technology and dynamic scheduling algorithms, has improved contract execution efficiency and effectively reduced costs. Despite significant achievements, there are still some shortcomings, such as the need to improve the accuracy of dynamic scheduling algorithms and the need to strengthen the universality of smart contract template mechanisms. In the future, optimization algorithms can be further improved to enhance the security and scalability of smart contracts, in order to meet the continuous development needs of the digital economy field.

## References

1. Agrawal TK, Angelis J, Khilji WA et al (2023) Demonstration of a blockchain-based framework using smart contracts for supply chain collaboration. *Int J Prod Res* 61(5):1497–1516
2. Lin SY, Zhang L, Li J et al (2022) A survey of application research based on blockchain smart contract. *Wireless Netw* 28(2):635–690
3. Fauziah Z, Latifah H, Omar X et al (2020) Application of blockchain technology in smart contracts: a systematic literature review. *Aptisi Trans Technopreneurship (ATT)* 2(2):160–166
4. Li T, Fang Y, Jian Z et al (2021) ATOM: Architectural support and optimization mechanism for smart contract fast update and execution in blockchain-based IoT. *IEEE Internet Things J* 9(11):7959–7971
5. Demertzis K, Iliadis L, Tziritas N et al (2020) Anomaly detection via blockchained deep learning smart contracts in industry 4.0. *Neu Comput Appl* 32(23):17361–17378
6. Leng J, Sha W, Lin Z et al (2023) Blockchained smart contract pyramid-driven multi-agent autonomous process control for resilient individualised manufacturing towards Industry 5.0. *Int J Prod Res* 61(13):4302–4321
7. Dolgui A, Ivanov D, Potryasaev S et al (2020) Blockchain-oriented dynamic modelling of smart contract design and execution in the supply chain. *Int J Prod Res* 58(7):2184–2199
8. Gao Z, Jiang L, Xia X et al (2020) Checking smart contracts with structural code embedding. *IEEE Trans Software Eng* 47(12):2874–2891
9. Zhang Y, Wang T, Yuen KV (2022) Construction site information decentralized management using blockchain and smart contracts. *Comput-Aided Civ Infrastruct Eng* 37(11):1450–1467
10. Wang W, Song J, Xu G et al (2020) Contractward: automated vulnerability detection models for ethereum smart contracts. *IEEE Trans Netw Sci Eng* 8(2):1133–1144
11. Kannengiesser N, Lins S, Sander C et al (2021) Challenges and common solutions in smart contract development. *IEEE Trans Software Eng* 48(11):4291–4318

# Tomato Disease Detection: Leveraging YOLOv8.2.0 for Accurate and Efficient Solutions



Hayder Mohammedqasim, Roa'a Mohammedqasem, Bilal A. Ozturk, Omar Akl, and Abdelkarim Boulahya

**Abstract** Tomato crops are very useful for our world, not just for many diets but also in the agricultural industry. Countries like Turkey are famous for quality tomatoes, which are used in the majority of dishes. Still, growing tomatoes is a difficult issue. The first reason is the diseases that affect the tomato plants and hence reduce the yield. Advanced deep learning techniques suggest that a novel technology is used to develop a system that can detect and categorize most of the three common diseases occurring to the tomato plants. In this work, the authors consider the use of the state-of-the-art technology, YOLOv8.2.0, which is a new release model, for the detection and categorization of the three most common diseases in tomato plants; this dataset solution worked with images and a dimension for tomato plants, which are infected by most common diseases. The achieved results are promising, in that the model achieved high accuracy for diseases such as blossom end rot, splitting, and sunscald disease. The accuracies are of 80.20%, 39.40%, and 70.10%, respectively, with an overall accuracy of 89.3% for this disease. It would, therefore, set the foundation of this research for an advanced system to be able to detect most of the diseases, with which crop yields—particular in Turkey—would be improved. Tomato crops are of tremendous importance in developing food security and economies worldwide; hence, they need to be protected from the diseases. Effective management of the diseases in question is still, nonetheless, complicated by technological challenges. This research makes a contribution through deep learning and advanced algorithms in the development of better disease detection systems—all the more indispensable in order to ensure tomato crop productivity and maintain a steady food supply.

**Keywords** YOLOv8 · Deep learning · Tomato diseases · CNN · Object detection

---

H. Mohammedqasim · R. Mohammedqasem · B. A. Ozturk · O. Akl (✉) · A. Boulahya  
Faculty of Engineering, Istanbul Aydin University, Istanbul, Turkey  
e-mail: [ioty989@gmail.com](mailto:ioty989@gmail.com)

R. Mohammedqasem  
e-mail: [rmoammedqasem@aydin.edu.tr](mailto:rmoammedqasem@aydin.edu.tr)

B. A. Ozturk  
e-mail: [bilalo@aydin.edu.tr](mailto:bilalo@aydin.edu.tr)

## 1 Introduction

Tomato crops occupy an important place in the agricultural sector and economy of Turkey. It provides high quality food and also has a major impact on the country food security [1]. This crop exports all around the world because turkey produce different type of tomato which quality are preum [2]. Seasons like Bursa, Izmir or Antalya has a high temperature and high productivity means weather and soils are perfect for tomato cultivation [3]. According to the stats are currently showing that tomato have a unique place in country area. For example, a following stats can show this. For example, adjusted estimated of total area of tomato production in Turkey in 2020 is 181,879 [4]. This huge place portraits how important tomato for Turkish agriculture and economic system.

**A Looming Threat** Tomato farming in Turkey is currently threatened by alarming rates of pathogen infestation, which affect both the plants and the fruit that they produce. Blossom end rot is a chief and very common problem of tomatoes that results in flat, brown or black sunken spots at the bottom of the fruit. Sunscald is a common issue as well, often due to sunlight. Sunburn can cause tomatoes to develop white or yellowish skin and tough flesh [4]. Cracked and split fruit is common, often due to uneven growth of the tomatoes. These diseases can not only decrease the yield of your tomatoes but also diminish their quality, rendering them less valuable in the market. Therefore, they are forced to use chemical treatments on their plantations to manage these issues—this in turn drives up the cost of production for farmers. This can make it very difficult for farmers to break even and continue farming in a sustainable manner. Because both land and water resources are scarce, these challenges must be addressed to increase food security in the face of Turkish farmer poverty [5]. Numerous techniques have been tested and automation has been used to identify sick tomatoes—the disease is indeed a lethal one. This strives to get as close as an exact number, however does not achieve the higher percentage goal of accuracy. In this paper, we resorted to AI for a remedy. We developed a model to accurately detect the three common diseases of tomato fruit by using deep learning algorithms. For this, we used the recently launched YOLOv8.2.0 and it has proven to have good results in terms of accuracy, precision, recall, using F1-score. For these three diseases, our model performed with the accuracy of 89.3%.

We have mostly followed a detailed approach. Therefore, in the dataset, we enriched an already existing dataset for the existence of most types of tomato diseases. We fine-tuned, enriched, and developed our model to achieve maximum performance from the task of fine-tuning. We trained the model using the state-of-the-art YOLOv8.2.0 algorithm, an advanced method in deep learning developed to identify three types of tomato diseases. We fine-tuned the model parameters for the best accuracy. We evaluated the developed model through the calculation of mean Average Precision, recall, and precision measurements to expose the strengths and reveal areas of the proposed model.

- In this study, we took a detailed approach to develop our model. The proposed model aims to enhance the existing Robflow dataset to include a variety of tomato diseases.
- Processing was used to clean and augment the Robflow dataset to make the model more robust and effective for disease prediction.
- YOLOv8.2.0 algorithm was used to detect three specific types of tomato diseases. Model parameters were fine-tune to obtain the best possible accuracy.
- Different evaluation metrics were used to evaluate the model such as mean Average Precision (mAP), recall, and precision to understand strengths and areas for improvement in the proposed model.

## 2 Related Work

Recent studies reveal that deep learning, particularly through advanced neural networks like CNNs, has greatly enhanced the detection of plant diseases. These techniques are excellent at analyzing detailed image features, identifying even the most complex disease symptoms in plants. By using large datasets of plant images, researchers have created highly accurate disease diagnosis systems. These systems can spot diseases early, enabling farmers to take quick action to prevent the spread and minimize yield loss. As a result, incorporating deep learning into plant disease detection is essential for advancing agriculture, boosting yields, and ensuring food security.

Nyarko et al. [6] employed a 15-layer CNN as the foundation for the SSD to increase detection of healthy tomato fruits and three classes of tomato fruit diseases. These results were attained by the CNN-SSD model introduced in the work in question and which outperformed other models by scores of detection precision. In general, the similar problem of tomato plant diseases detection and classification was solved using a Raspberry Pi by Rahul et al. [7] Both image processing and CNN-based classification were employed in achieving the desired results in diseased identification including late blight, gray spot, and bacterial canker. Another research was done by Phan et al. [8], YOLOv5m algorithm was used for segmentation and classification of the tomato fruits into ripe, immature, and damaged classes. The obtained models had very high prediction accuracy, moreover, in the prediction of ripened and non-ripened tomatoes that led to promising applications such as, automated tomato fruit harvesting. CAMYOLO, is one known as a developed and optimized version of YOLOv5 for detecting tomatoes, developed by Appe et al. [9] the proposed model incorporated an attentively calibrated Convolutional Block Attention Module (CBAM) that improved the accuracy of identifying small and overlapping fruits, particularly the tomatoes, with an average precision of 88.1%. This shows that, more research is still being made toward improving the identification of diseases on tomato fruits to increase knowledge and possibly bring improvement in the practices of growing tomatoes. Another research was made by Yang et al. [10] they developed a light weight tomatoes disease detection model that was built

using YOLOv8s, the test results show that the improved YOLOv8s network has a lower loss. Moreover, the proposed algorithm significantly reduced the model size from 22 to 16 M. Chen et al. [11] were able to improve a YOLOv5 model to accurately recognize plant diseases under complex natural conditions. During the experiments, sample images from the dataset were selected randomly to build training and testing sets. The test results showed that the improved YOLOv5 model accuracy was 70% which is 5.4% higher than the accuracy of the original YOLOv5 model. The precision values of detecting the diseases in the plants was 86.5% for the improved model, and 86.6% for the original model. Rajasree et al. [12] also conducted another study, they employed and enhanced YOLO-X model for tomatoes diseases detection. The technique introduced enhances the Spatial Pyramid Pooling layer, making it more effective at extracting valuable features from training data of different sizes. Researchers were able to increase the model's ability to identify wider range of diseases symptoms by combining variables from multiple layers and different sizes. The improved YOLO-X model achieved an improvement in test dataset accuracy and a 73.42% mean average precision on field-collected dataset. Hashim et al. [13] in another study utilized YOLOv3 to build a deep learning model to be able to detect plant diseases. After testing, the model's accuracy was rounding between 80 and 90%.

However, there is no prior work that provides a comparative study of different computer vision methods for tomato disease detection. Before the release of YOLOv8.2.0 [13], there have been studies on numerous deep learning networks that can be used for object detection. For instance, YOLOv8 presented some major advancements over prior versions: it boasts of high accuracy and speed. However, there is little detail regarding the use of YOLOv8.2.0 that is solely for identifying diseases in tomatoes.

### 3 Methodology

In this study, we followed a systematic approach (Fig. 1) to develop our model. First, we collected and enhanced an existing Robflow dataset, ensuring it had diverse examples of tomato diseases. To make the model even more robust and effective, we did data cleaning and augmentation as a preprocessing task. The YOLOv8.2.0 model is well known for its enhanced and advanced capabilities in deep learning for detecting the three specific types of diseases in tomato. We fine-tuned our model parameters [14] with a little more meticulous way to get the inference with the best possible output. We evaluated the model using metrics like precision, recall, and mean Average Precision, which give us an idea about its strengths and weaknesses.

**Fig. 1** Methodology flowchart





**Fig. 2** Sample of dataset images

The process of the identification of diseases in tomato fruits using the YOLOv8.2.0 algorithm is generally broken down as follows: data collection, data preparation, and model training and validation. This began by collecting various images showing both healthy and disease-affected tomatoes and then manually labeling them. The images, for making the dataset and, in turn, the model more robust, were resized followed by several data augmentation techniques. We have trained the YOLOv8.2.0, which can identify diseases accurately and fast enough. After that, it is extensively tested and validated by different metrics to validate the proposed real-world effectiveness of the model. This approach will allow for reliable development of a tomato disease detection system.

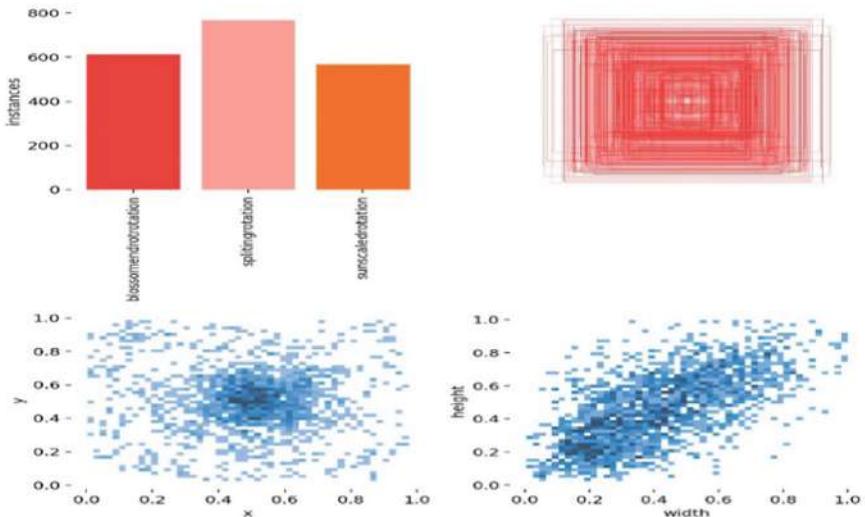
### 3.1 Data Description

The dataset [15] for the detection of tomato fruit disease contains 2158 images: 1878 images in the training set, 92 in the test set, and 188 in the validation set. The dataset contains images of a tomato for the diseases of blossom end rot, sunscald, and splitting. The bounding box of the disease region is introduced in the dataset. In addition, the data augmentation of rotation and flipping is performed to enhance the performance of the model. The annotations on this dataset are very detailed and important for training and testing on the YOLOv8.2.0 algorithm in classifying tomato fruit diseases (Figs. 2 and 3).

### 3.2 Architecture of YOLOv8.2.0

The architecture of YOLOv8.2.0 [16] has been categorized widely into three prime components:

- **Backbone:** This convolutional neural network undertakes feature extraction from an input image. The model chooses CSPDarknet53 for the backbone, while new cross-stage partial connections in many stages are deeply introduced through



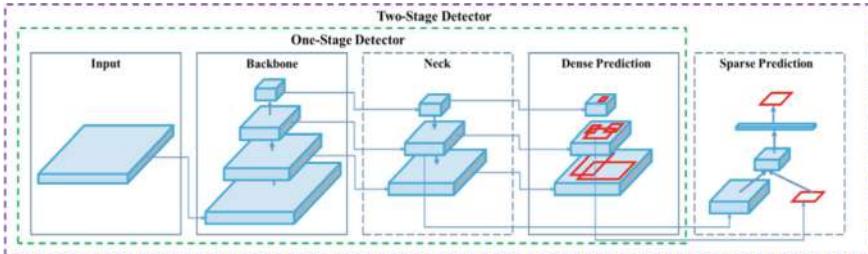
**Fig. 3** Data description

YOLOv8.2.0 for enhancement and bettering of the information flow quality and accuracy.

- **Neck:** The neck, as known as the feature extractor, integrates feature maps across different levels or stages of the backbone network to aggregate information at diverse scales. Unlike other YOLO versions, YOLOv8.2.0 does not use Feature Pyramid Network (FPN) to aggregate feature maps but instead use C2f module. This module combines the global semantic feature with the detailed spatial features, improving the detectability particularly for small objects.
- **Head:** The head acts as a decision-maker, where predictions have to be made. YOLOv8.2.0 utilizes several detection modules that consist of predicting the boxes of objects, the probability of an object being present, and the class probabilities of each grid cell in the obtained feature map. Each of these predictions is then accumulated to form the final detections as shown in the figure. Figure 4 illustrates the architecture of the YOLOv8.2.0 algorithm.

To prepare the dataset, we took crucial steps to enhance the images for training the YOLOv8.2.0 algorithm. This involved augmenting all images by rotating, flipping, and adjusting colors to train the model on different image variations. Additionally, we resized all images to 640 pixels by default. These preprocessing steps standardized and enriched the data, improving the model's ability to accurately diagnose tomato fruit diseases under various conditions (Fig. 5).

YOLOv8.2.0 was utilized as the foundational architecture for the proposed model due to its enhanced efficiency and flexibility [17], effectively addressing three key computer vision tasks [18]:



**Fig. 4** YOLOv8.2.0 architecture



**Fig. 5** Applying data augmentation on the data

- **Classification** [19]: In this task, the model must choose which class is most dominant in the given input image, and it returns two arrays of the class and the certainty level of the model. Classification is effective in determining whether or not image contains a part from a certain class.
- **Detection** [20]: The next level, detection, goes beyond classification, as it not only categorizes multiple objects in an image but also pinpoints the position of these objects within the image using axes.
- **Segmentation**: Unlike object detection, which is a simpler process whereby the system can determine if an object with certain features is present in an image or not, segmentation requires the system to identify which particular pixels in an image belong to which object, thus being more sophisticated. This technique as we have seen finds various applications and provides detailed information on objects in the image.

## 4 Results

We used Google Colab, a Python notebook environment popular among students and researchers, for tomato disease detection, as shown in Fig. 6. Google Colab seamlessly integrates with TensorFlow and PyTorch, offering users the choice of instance types like CPUs, GPUs, and TPUs, each with a 12-h runtime. To boost computational efficiency, we added a 12GB NVIDIA Tesla T4 GPU, which increased the complexity of our application at no extra cost. Initially, the training of the model was conducted for 100 epochs where 16 images at a time were taken from the database as a batch of images was used for efficient training but not compromising with the quality of the models. As another initiative, a new program coded in Python was also created, it uses OpenCV library [21] to assess the model and highlight the bounding boxes of the affected area that helps in better visualization of disease detection outcomes.

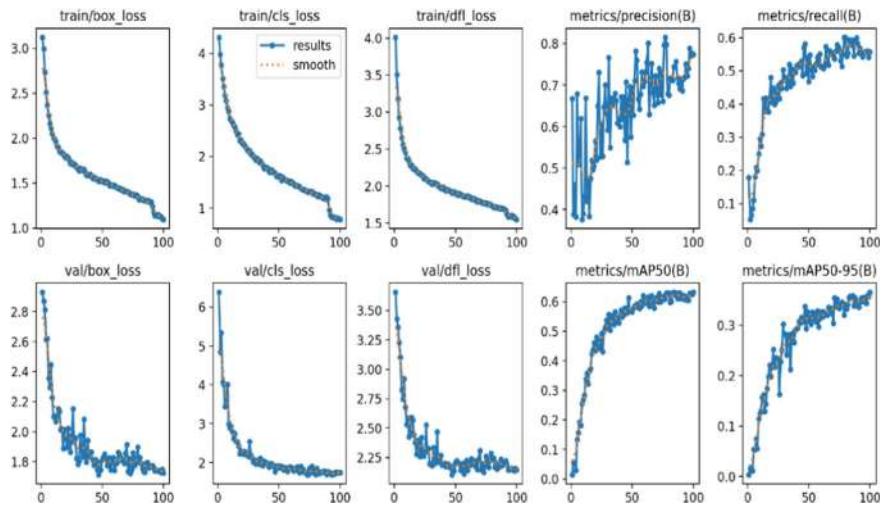
Figure 8 showcases how well the YOLOv8.2.0-based algorithm detects tomato diseases, using metrics like precision, recall, and mean Average Precision (mAP). Precision reflects the algorithm's high accuracy and low false alarm rate, showing the proportion of true positives in each detection. A high recall curve indicates the algorithm's effectiveness in identifying diseased plants accurately. The mAP offers a thorough evaluation by combining both accuracy and recall across different diseases. These metrics help ensure the model avoids both over-diagnosis and under-diagnosis, providing a balanced and reliable detection system (Fig. 7).

Figure 8a shows how the model's accuracy changes with different confidence levels. Generally, higher confidence means higher precision, indicating the model's ability to distinguish well between high and low-confidence detections. Figure 8b reveals that recall values are consistently higher than precision but lower than those for binary classification problems, highlighting the true positives at various confidence levels. Initially, the model identifies many diseases with low probabilities, showing high sensitivity, but it becomes stricter as confidence increases. Figure 8c illustrates the trade-off between accuracy and recall, aiming to detect as many diseases as possible while minimizing false predictions. Figure 8d uses the F1-confidence curve to evaluate performance, with the peak showing the best balance between precision and recall for the dataset.

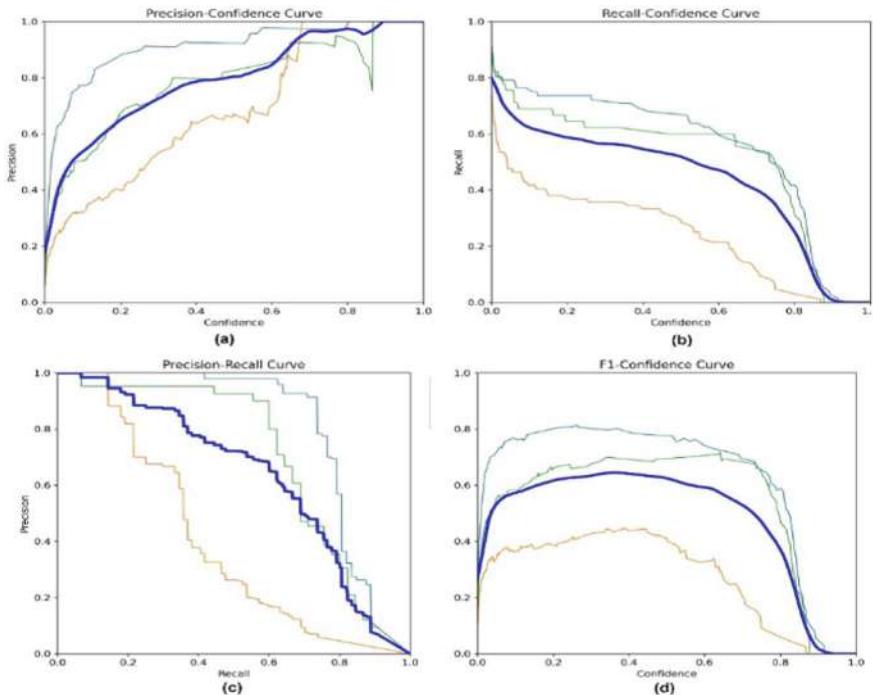
A confusion matrix has been provided in Fig. 9 depicting the various types of diseases identified by this model. Each cell contains the probability of the model making an accurate prognostication or misdiagnosing a particular disease for another. Summing up, it should be pointed out that in the case of the discussed model of scoring

**Fig. 6** Sample from experimental results

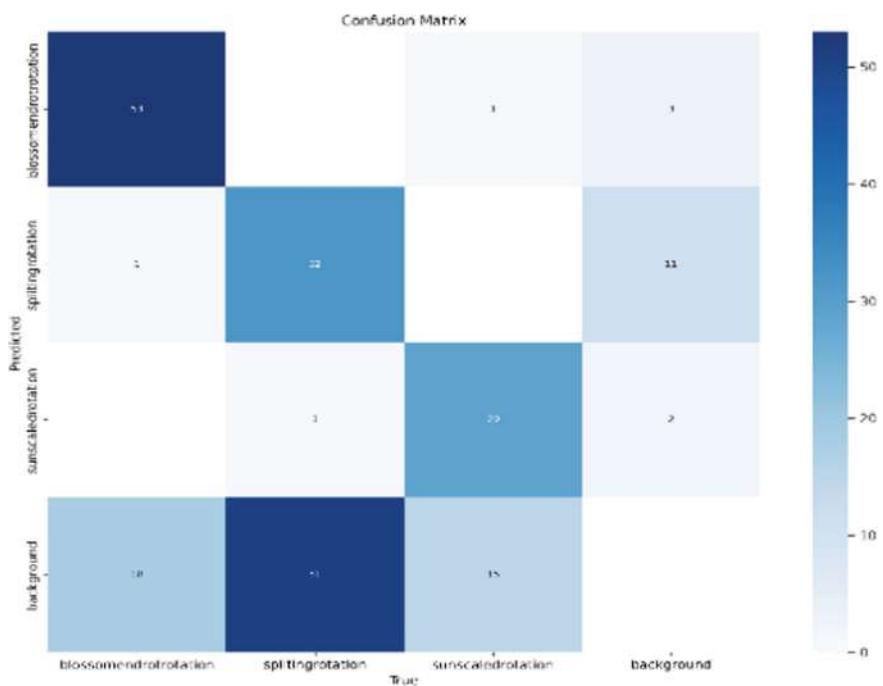




**Fig. 7** Effectiveness of the algorithm



**Fig. 8** **a** Precision-Confidence Curve—**b** Recall Confidence Curve—**c** Precision-Recall Curve—**d** F1 Confidence Curve



**Fig. 9** Normalized Confusion Matrix

the tester's performance, the obtained accuracy rate was as high as 89.3% across all diseases. However, these statistics merely reflect average performance which does not give information about variations of performance in the case of each separate disease.

In Table 1, the proposed model is compared with the model that was proposed by a previous study (Table 2).

The final results are presented in the table. Each disease has been detected with different accuracies. After being compared with the results of the previous study, our model shows that it is more efficient and more accurate (Table 3) with an overall accuracy of 89.3%.

Rotation have been applied to the images of the dataset, which is a type of data augmentation techniques in order to enhance the dataset. Table 3 shows the precision, recall, and the accuracy of each class after being rotated.

**Table 1** Comparing between our model and previous study

Aspect	Recent study [21]	Our model
Algorithm	YOLOv8	YOLOv8.2.0
Dataset	Balanced data Computer Vision Project	Balanced data Computer Vision Project
Annotation	Bounding boxes for various plant diseases	Bounding boxes for various plant diseases
Data augmentation	Applied (rotation, flipping, color adjustments, image resizing)	Applied (rotation, flipping, color adjustments, image resizing)
Preprocessing	Resized images to 640 pixels (By default)	Resized images to 640 pixels (By default)
Hardware	Google Colab with 12GB NVIDIA Tesla T4 GPU	Google Colab with 12GB NVIDIA Tesla T4 GPU
Training duration	50 Epochs	100 Epochs
Batch size	3	16
Performance metrics	Precision, Recall, F1-score, mAP	Precision, Recall, F1-score, mAP

**Table 2** Comparison table between the results of our proposed model with the results of a previous study

Study	Disease type	Precision (%)	Recall (%)	F1-score (%)	mAP
Zayani et al. [21]	Blossom end rot	86.44	70.84	80.0	0.59
	Sunscald	85.71	66.67	70.0	0.61
	Splitting	74.07	90.90	50.0	0.70
	Overall	66.0	60	65	0.66
Proposed	Blossom end rot	91.2	87.5	89.3	0.802
	Sunscald	88.5	86.9	86.7	0.701
	Splitting	90.0	88.2	89.1	0.694
	Overall	89.9	86.9	88.4	0.79

**Table 3** Accuracy metrics of the rotated classes

	Blossom end rot rotation (%)	Splitting rotation (%)	Sunscald rotation (%)
Precision	89	65	79
Recall	98	38	95
mAP	80.20	39.40	70.10

## 5 Limitations

The model we created is pretty accurate, but it does have some downsides. It works better for some diseases than others; for example, it's more accurate at spotting blossom end rot (80.2%) and sunscald (70.1%) than splitting (39.4%). This shows

that the model has more trouble with certain diseases, which could make it less reliable overall. Also, the data we used to train the model might not cover all the different environmental conditions and tomato variations found in different regions. This could affect how well the model performs in real-life situations where conditions are different from what it was trained on.

## 6 Conclusion

Tomato diseases cause huge losses to farmers and the agricultural economy in general. They lead to reduced crop yield and quality, financial strains on the farmers, and a rise in prices for the consumers. This disease also increases the cost of purchasing fungicides for the farmer. This study recommends that deep learning through the YOLOv8.2.0 approach be conducted to help farmers in their identification of the three most common diseases affecting their tomato plants. The proposed model estimated a site-recorded accuracy of 89.3% classification. That will be an appropriate tool for a farmer in diagnosing diseases on time, thus acting promptly to rescue the crops. Future research could make the model more accurate and reliable by increasing datasets in terms of diseases, symptoms, and environmental parameters. Exploring different deep learning techniques could also further improve performance. Developing a user-friendly app for field use and conducting real-world tests would also be beneficial in assessing the model's practical effectiveness.

## References

1. Kayikcioglu HH, Duman İ, Asciogul TK, Bozokalfa MK, Elmacı ÖL (2020) Effects of tomato-based rotations with diversified pre-planting on soil health in the Mediterranean soils of Western Turkey. Agric Ecosyst Environ 299:106986. <https://doi.org/10.1016/J.AGEE.2020.106986>
2. Arslan Ş, Arisoy H, Karakayaci Z (2022) The Situation of regional concentration of tomato foreign trade in Turkey. Turkish J Agric Food Sci Tech 10(2):280–289. <https://doi.org/10.24925/TURJAF.V10I2.280-289.4767>
3. Şalli B, Kavlak B, Sunar AF (2024) Improving multi-crop area assessment through Bootstrapping: a focus on tomato fields. Remote Sens Appl 33:101115. <https://doi.org/10.1016/J.RSASE.2023.101115>
4. Bilgili A (2023) Thermal image processing for automatic detection of fusarium root and crown rot disease in tomato plants. Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi 14(4):611–619. <https://doi.org/10.24012/DUMF.1340922>
5. Türkten H, Ceyhan V (2023) Environmental efficiency in greenhouse tomato production using soilless farming technology. J Clean Prod 398:136482. <https://doi.org/10.1016/J.JCLEPRO.2023.136482>
6. Nyarko BNE, Bin W, Jinzhi Z, Odooom J (2023) Tomato fruit disease detection based on improved single shot detection algorithm. J Plant Prot Res 63(4):405–417. <https://doi.org/10.24425/JPPR.2023.146877>
7. Rahul J, Sharma LD, Bhardwaj R, Singh RS (2023) Disease detection in tomato leaves using raspberry pi-based machine learning model. Lect Notes Electr Eng 982:425–434. [https://doi.org/10.1007/978-981-19-8136-4\\_35](https://doi.org/10.1007/978-981-19-8136-4_35)

8. Phan QH, Nguyen VT, Lien CH, Duong TP, Hou MTK, Le NB, Classification of tomato fruit using Yolov5 and convolutional neural network models. *Plants* 12(4):790. <https://doi.org/10.3390/PLANTS12040790>
9. Appe SN, Arulselvi G, Balaji GN (2023) CAM-YOLO: tomato detection and classification based on improved YOLOv5 using combining attention mechanism. *PeerJ Comput Sci* 9:e1463. <https://doi.org/10.7717/PEERJ-CS.1463>
10. Yang G, Wang J, Nie Z, Yang H, Yu S (2023) A lightweight YOLOv8 tomato detection algorithm combining feature enhancement and attention. *Agronomy* 13(7):1824. <https://doi.org/10.3390/AGRONOMY13071824>
11. Chen Z et al (2022) Plant disease recognition model based on improved YOLOv5. *Agronomy* 12(2):365. <https://doi.org/10.3390/AGRONOMY12020365>
12. Beulah Christalin Latha C (2024) Improved YOLO-X model for tomato disease severity detection using field dataset. *IJACSA* Int J Adv Comput Sci Appl 14(9):2023, Accessed: May 27, 2024. [Online]. Available: [www.ijacsathe.org](http://www.ijacsathe.org)
13. Chairma Lakshmi KR, Praveena B, Sahaana G, Nithya Jenev J, Gnanasekaran T, Hashim M (2023) Yolo for detecting plant diseases. In: Proceedings of the 3rd international conference on artificial intelligence and smart energy, ICAIS 2023, pp 1029–1034. <https://doi.org/10.1109/ICAI56108.2023.10073875>
14. Mohammedqasim H, Ahmed Jasim A, Mohammedqasem A, Ata O (2024) Enhancing predictive performance in covid-19 healthcare datasets: a case study based on hyper Adasyn Over-Sampling and genetic feature selection. *J Eng Sci Technol* 19(2):598–617
15. “balanceddata - v3 2023-10-24 12:04am.” Accessed: May 27, 2024. [Online]. Available: <https://universe.roboflow.com/research-proj-mgap9/balanceddata-y4ox0/dataset/3>
16. Barlybayev A et al (2024) Personal protective equipment detection using YOLOv8 architecture on object detection benchmark datasets: a comparative study. *Cogent Eng* 11(1). <https://doi.org/10.1080/23311916.2024.2333209>
17. Terven J, Córdova-Esparza DM, Romero-González JA (2023) A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS. *Mach Learn Knowl Extract* 5(4):1680–1716. <https://doi.org/10.3390/MAKE5040083>
18. Jasim AA, Hazim LR, Mohammedqasim H, Mohammedqasem R, Ata O, Salman OH (2024) e-Diagnostic system for diabetes disease prediction on an IoMT environment-based hyper AdaBoost machine learning model. *J Supercomput* 1–26. <https://doi.org/10.1007/S11227-024-06082-0/TABLES/4>
19. Mohammedqasem R et al (2023) Multi-objective deep learning framework for COVID-19 dataset problems. *J King Saud Univ Sci* 35(3):102527. <https://doi.org/10.1016/J.JKSUS.2022.102527>
20. Mohammedqasem R, Mohammedqasim H, Ata O (2022) Real-time data of COVID-19 detection with IoT sensor tracking using artificial neural network. *Comput Electr Eng* 100:107971. <https://doi.org/10.1016/J.COMPELECENG.2022.107971>
21. Zayani HM et al (2024) Deep learning for tomato disease detection with YOLOv8. *Engi Technol Appl Sci Res* 14(2):13584–13591. <https://doi.org/10.48084/ETASR.7064>

# Data-Driven Facial Image Synthesis from Text Descriptions with Deep Fusion GANs



Naveen Ananda Kumar Joseph Annaiah  and Mohan Mahanty 

**Abstract** In crime investigation, the need to generate accurate facial images from textual descriptions is crucial for identifying suspects and solving cases. Traditional approaches consume more time and accuracy also very low. In recent times, researches using deep learning-based GAN architectures, such as DCGAN and StackGAN, have shown promise in generating realistic images. However, these models often consist of multiple generators, which can introduce ambiguity and complexity in the generation process, particularly when dealing with textual descriptions. So, we proposed a Deep Fusion Generative Adversarial Network architecture to overcome the disadvantages associated with the existing systems. Our model integrates seven new layers termed as UP Blocks to enhance feature extraction and synthesis, while also incorporating a discriminator with matching-sensitive gradient regularization (MS-GR) to improve the discrimination between real and generated images. Through extensive experimentation, we demonstrate the effectiveness of our methodology in producing high-quality facial images that closely align with the provided textual descriptions. For evaluation of the modal, we used Frechet Inception Distance and inception score as metrics and our model achieves the inception score of  $1.318 \pm -0.225$  and Frechet Inception Distance score of 30.45, which surpasses the existing models.

**Keywords** GAN · Text-to-image · Deep Fusion GAN · CelebA dataset

---

N. A. K. J. Annaiah (✉)  
Tekinvaderz LLC, Florida, USA  
e-mail: [naveenjannaiah@gmail.com](mailto:naveenjannaiah@gmail.com)

M. Mahanty  
Department of Computer Science and Engineering, Vignan's Institute of Information Technology,  
Duvvada, Visakhapatnam, Andhra Pradesh, India

## 1 Introduction

Synthesizing realistic images from textual descriptions gives an impressive undertaking inside the realm of Deep Learning. The primary aim is to create pictures that faithfully represent the enter descriptions, executed thru the utility of Generative Adversarial Networks (GANs). While present efforts have in large part focused on generating simplistic pics like vegetation or birds from captions, this work ventures into the world of Text-to-Face technology (T2F), the subset of Text-to-Image (T2I) generation. The potential packages span various domain names which include Forensic Science, Animation, Digital Marketing, and Art.

Within the realm of artificial intelligence, Generative Adversarial Networks (GANs) distinguish themselves as sophisticated models that fall under the category of Deep Learning. Deep Learning consists of complex neural networks, emulates the functioning of the human brain to discern patterns in large unstructured datasets. This technique diverges from conventional machine learning algorithms, which rely on established data to generate predictions and discover patterns. This paper aims to utilize the power of Deep Learning, particularly through Generative Adversarial Networks (GANs), to expand the capabilities of photo synthesis. This will open up new possibilities for innovative applications in numerous industries.

Generative Adversarial Networks (GANs) have gained prominence due to their remarkable capacity to rapidly generate realistic images, establishing new benchmarks for efficiency and authenticity in image generation. Generative Adversarial Networks (GANs) [1] have greatly improved the ability to create very authentic face photos, which can be misused to create fraudulent social media profiles and propagate misinformation, potentially leading to serious consequences. GANs provide many techniques for generating images that closely resemble genuine ones with exceptional quality. The conversion of written descriptions into facial images has significant potential in multiple fields, including entertainment, virtual reality, and assisting in activities like finding missing individuals and resolving criminal cases.

Art comprises a diverse range of human activities and results, frequently showcasing creative talent and technical expertise, eliciting emotions, or conveying abstract concepts. During criminal investigations, law enforcement organizations frequently enlist the aid of artists to create sketches of suspects using verbal descriptions. However, utilizing this traditional method can be time-consuming and result in substantial delays in the conclusion of criminal cases [2].

Utilizing Generative Adversarial Networks (GANs) for Text-to-Image (T2I) generation entails using textual descriptions, specifically facial traits retrieved through Natural Language Processing (NLP), to create lifelike images. Natural Language Processing (NLP) is an advanced form of machine learning that allows computers to accurately and precisely analyze, manipulate, and comprehend human language [3]. The BERT model can be utilized as an input encoder. BERT, which stands for Bidirectional Encoder Representations from Transformers, is a highly advanced Deep Learning language model designed to improve the efficiency and speed of various Natural Language Processing (NLP) activities [4].

## 1.1 Generative Adversarial Networks

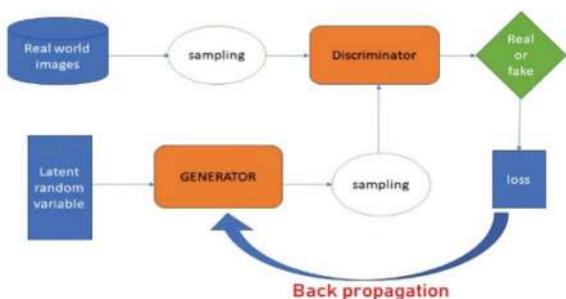
GANs utilize a distinctive training process, as seen in Fig. 1, which involves emulating a supervised learning strategy to address generative modeling tasks. GANs consist of two competing sub-models, namely the Generator and the Discriminator, which are both integral parts of its architecture. The Generator, a component of a neural network, is tasked with creating data instances, whereas the Discriminator's function is to determine their legitimacy. The Discriminator plays a crucial role in improving the authenticity of generated outputs by determining if a data instance is genuine or manufactured.

In GANs, the generator model strives to trick the discriminator by continuously improving its output to appear progressively more believable. This iterative procedure entails combining backpropagation with a competitive interaction between two networks, namely the Generative Network (G) and the Discriminative Network (D). While G generates artificial images, D evaluates and categorizes them as either real or artificial, contributing to the refinement of the generative process.

The generator takes the random noise as input and tries to mimic the training dataset to generate fake images and it aims to generate new samples that resemble the real data. In its role as a binary classifier, the discriminator distinguishes between genuine data samples sourced from the training dataset and counterfeit samples crafted by the generator. Simultaneously, the discriminator endeavors to accurately classify both authentic and synthetic samples, contributing to the iterative refinement process of the GAN. While the discriminator tries to correctly classify real and fake samples. If the discriminator loss is high, the generated image is backpropagated to the generator to create more quality phony images, and this process will be continued until the Discriminator fails to identify the generated image as real or fake (Fig. 1).

There are different variants of GAN, namely Conditional GAN, Deep Convolutional GAN, Variational autoencoder GAN, Self-attention GAN, Transformer GAN, Bidirectional GAN, Cycle GAN, Flow-GAN, and versions of style GAN. Existing GAN models, including DCGAN, FTGAN, StyleGAN, and StackGAN, have demonstrated the ability to achieve a remarkable accuracy of 57% similarity with real-time images [5]. In this paper, we used the proposed Deep Fusion GAN (DFGAN),

**Fig. 1** GAN training process



by considering the instability observed in the training process of previous FTGAN frameworks [6].

We can give input as facial attributes like some of attributes are “mouth slightly open,” “smiling,” “gray hair,” “no beard,” “bags under eyes,” “bushy eye brows,” “heavy makeup,” “oval face,” etc. In GANs, facial attributes such as age, gender, hair color, and facial expression are transformed into vectors and given as input to the GAN model. The Generator utilizes vectors obtained from datasets such as CelebA to generate randomized facial images. This approach frequently necessitates a greater number of iterations in comparison with previous GAN models in order to generate high-quality, naturalistic images.

## 2 Literature Review

Anukriti Kumar et al. conducted a study on creating realistic facial images based on written descriptions using a sketch refinement technique [7]. They used the CelebA dataset, where all the images have a high resolution of  $256 \times 256$  pixels. This played a crucial part in achieving significant improvements in the results. The authors introduced the StackGAN architecture, which utilizes a dual-stage procedure to generate images with a wide range of facial expressions, including broad smiles or expressions of sadness. The model demonstrated impressive performance, obtaining a noteworthy inception score of  $4.04 \pm 0.05$  over 10 Epochs.

Xiang Chen and their collaborators carried out study using the FTGAN model [6] to generate lifelike human faces based on textual descriptions of facial features. Significantly, the image and text encoders are trained concurrently. The FTGAN model is specifically built to produce images at three distinct scales, spanning from low resolution (e.g.,  $4 \times 4$  quality) to high resolution (e.g.,  $64 \times 64$  quality). Authors claimed that their model exhibits substantial improvement in generated image quality, reaching a similarity of 59% with ground truth images, and getting an inception score of  $4.61 \pm 0.05$  for the CUB dataset.

Through simultaneous training of image and text encoders, M. Zeeshan Khan et al. concentrated their research on improving GANs [8] for producing realistic images. They merged two different datasets from CelebA and LFW in order to augment the quality of the images. As an output, their model significantly generated two similar images that corresponded to the same input descriptions, each with a resolution of  $256 \times 256$ . The achieved quality was evidenced by FSD score of 1.218 and FID score of 44.62.

Osaïd Rehman Nasir et al. and colleagues leveraged finely grained textual descriptions to generate facial images [9]. The method they utilized encompassed the utilization of an algorithm to produce captions corresponding to images contained within the CelebA dataset. For multimodality assistance, they combined the use of DCGAN and GANCLS loss. They used a strategy that involved injecting noise and flipping

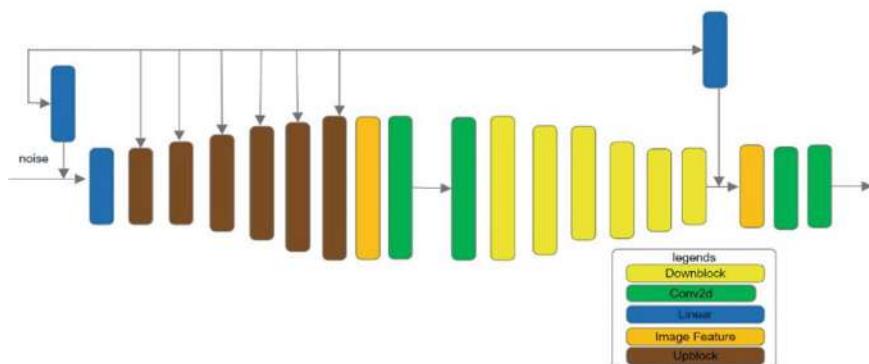
labels between genuine and false images in order to improve discriminator performance. Although their approach achieved an inception score of  $1.41 \pm 0.78$ , it was not particularly evaluated on the CelebA dataset.

Kushal Jivarajani and his colleagues developed a technique using Generative Adversarial Networks (GANs) to automatically create realistic human faces based on text descriptions [10]. Their approach involved training a VQGAN model on the CelebA dataset, while simultaneously pretraining the text using a CLIP conditioning model. Authors claimed that their model demonstrated much-improved precision and efficiency in the process of associating text with images. Nevertheless, the system's efficiency is hindered by slower processing on less powerful devices and a need for more detailed feature descriptions for human face images.

### 3 Proposed Model

In our study, we proposed the Deep Fusion GAN (DFGAN) model [11], consists of three essential components: (i) Input Encoder, (ii) Generator, and (iii) Discriminator, as illustrated in Fig. 2. Every component holds significant importance in the process of synthesizing images based on textual descriptions. Unlike conventional text encoders that directly convert input textual descriptions into semantic vectors, our proposed approach involves a preprocessing step. Before sentence encoding, utilizing a well-established algorithm, we generate captions by leveraging the attributes embedded within the CelebA dataset. This distinctive preprocessing phase is designed to enrich the semantic characteristics of the produced images during the training of the model, thereby potentially elevating the accuracy and intricacy of the synthesized images.

This framework enables high-resolution image generation through a single pair of generator and discriminator, while incorporating text information and visual feature maps via. Within the DF-GAN architecture, a generator, discriminator, and



**Fig. 2** Deep Fusion GAN

multiple pre-trained text Deep Text-Image Fusion Blocks (DF Blocks) embedded within UP Blocks are introduced. The model exhibits exceptional competence in producing genuine images that closely correspond to the provided textual descriptions by employing the Matching-Aware Gradient Penalty (MA-GP) technique and implementing a one-way output strategy.

### **3.1 BERT**

We utilized sentence-transformers model (Sentence BERT) to furnish the generator with a semantic vector that encapsulates the input sentence [12]. This model is designed to convert phrases and paragraphs into a 768-dimensional dense vector space. It has a wide range of features, including clustering and semantic search, which makes it highly useful for various applications. Sentence BERT is a modified version of BERT that is specifically designed to generate sentence embeddings that carry semantic meaning. It is highly successful for various Natural Language Processing tasks [4].

The authors showcased that conventional approaches for obtaining sentence embeddings using BERT fell short in achieving satisfactory results, particularly in tasks like textual similarity assessment. The architecture of Sentence BERT relies heavily on the training data at hand. Throughout our experimentation, we delved into diverse network structures and objective functions, aiming to enhance performance across different benchmarks.

#### **Regression Function**

In this function, cosine similarity between the embeddings of two sentences can be calculated as in Fig. 3. Mean squared error loss serves as the principal objective function guiding our computational processes.

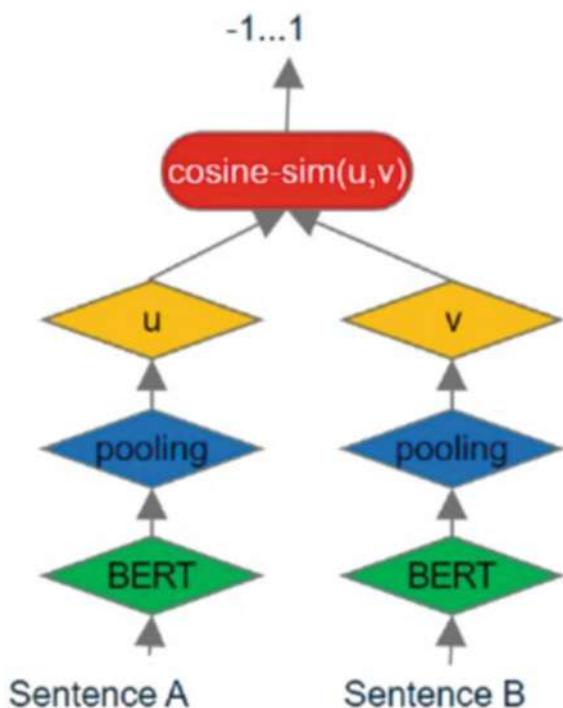
#### **Classification Function**

We combine the sentence embeddings  $u$  and  $v$  by joining them together and including the element-wise difference  $|u - v|$ . This combined result is then multiplied by the trainable weight  $W_t \in R$ , which belongs to the set of real numbers  $3n \times k$  as shown Fig. 4.

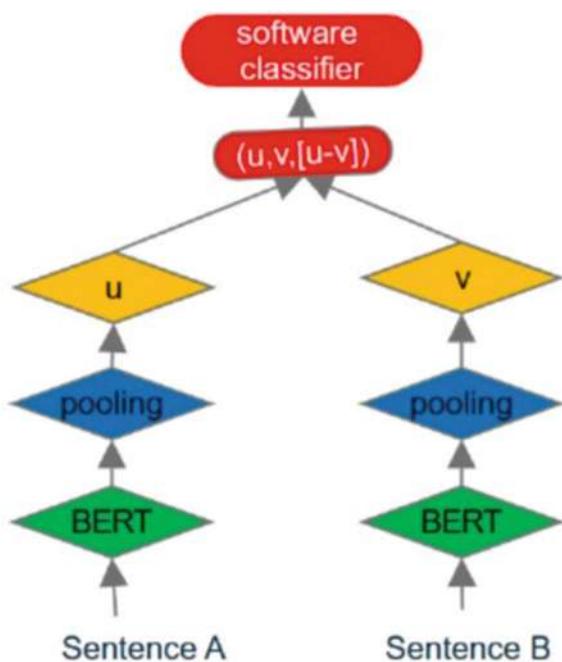
### **3.2 Generator**

Due to the inherent instability observed in GAN models, earlier text-to-image GANs often employed stacked architectures to generate high-resolution images from lower-resolution inputs. Nonetheless, the stacking of multiple generators and discriminators may introduce entanglements among different components, potentially yielding final refined images that merely resemble a blend of indistinct shapes.

**Fig. 3** SBERT regression function [12]



**Fig. 4** SBERT classification function [12]



Inspired by the methodology delineated in sDF-GAN [11], our approach diverges from the conventional stack architecture framework. Instead, we opt for a singular generator equipped with additional layers, this facilitates the direct generating of high-resolution images from noise vector. With a focus on our generator, it works with multiple inputs: A sentence vector which is encoded by our text encoder and the noise vector sampled by Gaussian distribution, ensuring diversity in the generated images. Initially, the noise vector undergoes preprocessing via a fully connected layer and subsequent reshaping. Following this, a sequence of UP Blocks is utilized to gradually up-sample the image features. Each UP Block includes an up-sampling layer, a residual block, and DF Blocks, enabling smooth integration of text and image characteristics during the image generation procedure. Ultimately, a convolutional layer is utilized to convert the image features into concrete images.

### **UP Block**

Within our DF-GAN’s generator, we integrate 7 UP Blocks, each housing multiple fusion blocks to optimize textual information utilization during fusion. Leveraging the established Deep Fusion Block (DF Block) [11], to augment fusion capabilities, we integrate a set of Affine transformations and RELU layers within our framework.

### **Down Block**

The DF Block draws its inspiration from Conditional Batch Normalization (CBN) [13] and Adaptive Instance Normalize (ADAIN) [14] and both of which integrate the Affine transformation [15]. However, whereas CBN and ADAIN incorporate normalization layers to align feature maps toward a normal distribution, the normalization process may contradict the goal of the Affine Transformation, which aims to amplify the distinctions among various samples. Consequently, this normalization step is omitted as it proves counterproductive for the conditional generation process.

Furthermore, the depth of our Deep Fusion Block enhances the text to image fusion process, enriching its capabilities. We incorporate multiple Affine layers stacked together, with a RELU layer interspersed between them. This strategy fosters the diversification of graphical features and expands the interpretation spaces, thereby accommodating a broader range of visual features corresponding to distinct text descriptions.

### **3.3 Discriminator**

The discriminator undertakes image processing through a sequence of Down Blocks, transforming them into image features. Afterward, the sentence vector is replicated and combined with the image features. Following this integration, an adversarial loss is calculated to assess both the realism and semantic coherence of the inputs. Through distinguishing between generated images and authentic instances, the discriminator motivates the generator to produce images of higher quality and improved semantic

coherence between text and image features. Drawing inspiration from this discriminator integrates, matching-sensitive gradient regularization (MS-GR) and one-way output mechanism aims to steer the generator toward generating images that exhibit both heightened realism and enhanced text to image semantic coherence.

### Matching-Sensitive Gradient Regularization

In this section, we embark on a comprehensive exploration of the unconditional gradient penalty [16], offering novel insights and perspectives. Following this, we advance to further elaborate on this notion, introducing the innovative matching-sensitive gradient regularization (MS-GR) meticulously crafted to elevate the interpretive coherence of text and images within the domain of text to image creation.

Based on our earlier examination, we infer that implementing gradient penalties on target data facilitates the creation of a more favorable loss landscape, thereby aiding the generator's convergence. This observation holds particular significance in the realm of text to image creation. The discriminator in text to image creation processes four types of inputs. To ensure semantic coherence between text and visual components, our emphasis is on applying gradient penalties to real data paired with matching text—essentially, the target data for text-to-image compilation. Consequently, within framework of MS-GR, the gradient penalty is specifically enforced on real images accompanied by matching text.

Through the integration of the MS-GR loss as regularization method within the discriminator, the model showcases enhanced convergence toward real data that harmonizes seamlessly with the provided textual context, thereby yielding generated images that closely mirror the textual descriptions. Furthermore, as the discriminator undergoes joint training within our network architecture, it effectively deters the generator from producing adversarial attributes akin to those distinct, permanent auxiliary networks. Moreover, MS-GR provides the added advantage of dispensing with the need for extra networks to ensure text to image consistency. Given that gradients are evaluated via the backpropagation procedure, the only additional computation required is the summation of gradients, rendering it computationally more efficient compared to the use of supplementary networks.

### One Way Output

In prior text-to-image GANs, such as those referenced in [17, 18]. One pathway determines the authenticity of the image, while the other combines the image features and vector to assess text to image semantic stability. Where the two path way approach undermines the efficiency of MS-GR and impedes generator's convergence rate. Specifically, the conditional deficit generates the gradient  $\alpha$  that points toward real and corresponding inputs following backpropagation, while the unconditional deficit yields only gradient  $\beta$  directed solely at real images. However, the resultant gradient, being merely the sum of  $\alpha$  and  $\beta$ , fails to accurately guide toward real and corresponding data points as intended. This deviation in the final gradient, given the generator's objective of producing real and text-matching images, falls short of achieving

optimal text to image semantic consistency and decelerates the generator's cohesion. Hence, we opt for the one path way approach [11] in text to image analysis, as proposed in to address these shortcomings.

## 4 Results and Analysis

### 4.1 Dataset

Our model utilizes the CelebAFaces Attributes dataset (CelebA) [19], which comprises 202,599 face images sized  $178 \times 218$ , featuring various celebrities. This dataset encompasses 10,177 distinct identity faces and includes 40 binary attribute annotations per image, such as arched eyebrows, attractiveness, presence of bags under the eyes, baldness, and more. Each attribute is assigned a value of 1 or  $-1$ , denoting its presence or absence in the image, respectively. Additionally, the celebA dataset provides height and width information for each image, with all images stored in JPG format.

### 4.2 Evaluation Metrics

To quantitatively assess the performance of the proposed model and to facilitate a rigorous comparison with existing state-of-the-art(SOTA) models, we used the evaluation metrics such as the Fréchet Inception Distance (FID) and the Inception Score (IS).

#### Frechet Inception Distance

The FID serves as a metric to gauge the realism and diversity of images produced by GANs. Realism refers to the extent to which generated images resemble real ones, particularly in the context of human subjects. Diversity, pertains to the degree of variation among generated images, rendering them intriguing and innovative. FID is instrumental in evaluating individual images generated by GANs, analyzing the impact of modifications in neural network models on realism, and comparing the efficacy of various GAN models in image generation tasks. It effectively captures both visual quality and diversity within a single metric. A less FID score indicates a closer resemblance between generated and truth images, aiding in identifying anomalies such as additional fingers or misplaced facial features. The FID score is measured by Eq. 1.

$$d^2 = |(|\mu_1 - \mu_2|)|^2 + \text{Tr}\left(C_1 + C_2 - 2 * \sqrt{C_1} * C_2\right) \quad (1)$$

### Inception Score (IS)

IS defines for widely used metric for evaluating the images produced by GANs. This metric quantifies the realism of a GAN's output, encompassing two crucial aspects: The variety of generated outputs and the perceptual quality of each individual image. A high IS signifies that the generated images exhibit both variety and clarity, while a low score indicates deficiencies in either or both of these aspects. Therefore, a higher IS indicates the GAN's capability to produce a broad spectrum of distinct and recognizable images. As shown in Eq. (2),  $p(y|x)$  is a conditional probability of every image.

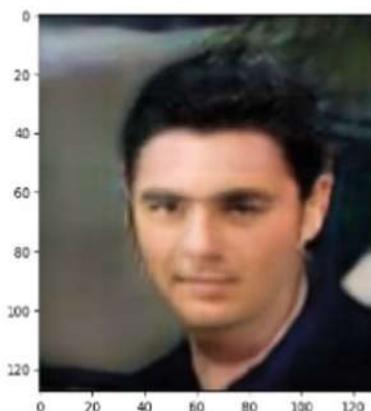
$$\text{KL} = p(y|x) * (\log(p(y|x)) - \log(p(y))) \quad (2)$$

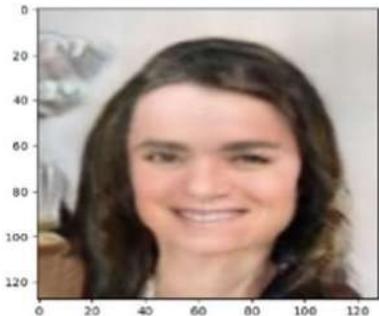
### 4.3 Results

We evaluated our model by submitting the textual descriptions. When we submitted the description as “He has a 5 o' clock shadow. His hair is black and straight. He has big lips, a big nose, bushy eyebrows and a pointy nose. The man seems attractive and young.” The proposed DFGAN successfully generates a human face as shown in Fig. 5 that closely resembles to the given textual description. The generated human face is youthful and attractive, reflecting the model's capability to convert subtle textual hints into realistic facial features.

When we prompted the textual description “The lady has pretty high cheekbones. She has brown hair. She has a big nose and a slightly open mouth. She is smiling and looks young,” the proposed DFGAN model generates a human female face as shown in Fig. 6. The generated image is consistent with the given textual description. The

**Fig. 5** Generated image 1



**Fig. 6** Generated image 2**Table 1** IS and FID score calculated of different GANs

Model	Inception score(IS)	FID score
Attn GAN	$1.062 \pm 0.051$	41.73
Stack GAN	–	46.02
DF GAN (Proposed model)	$1.138 \pm 0.225$	30.45

obtained results prove that the model can convert the given textual features into a visually consistent and realistic human face.

When we compared the performance of our model to the existing models, our model attains better IS, FID scores compared to existing state of the art models as shown in Table 1.

Generation of the facial images from the text descriptions is a typical task, because the complex structure of the human faces and the similarities between the faces may results some wrong predictions. So, there a scope for improvement in the field of research.

## 5 Conclusion

The research focuses on generating highly realistic facial images from textual descriptions to aid criminal investigations. A novel Deep Fusion Generative Adversarial Network (DF-GAN) is proposed to generate highly realistic facial images from textual descriptions. By integrating a BERT model, DF-GAN effectively aligns textual and visual information. The model's architecture enables direct generation of high-resolution images and incorporates advanced discriminator techniques for improved image-text coherence. Experimental results demonstrate superior performance compared to existing methods. Future research will focus on enhancing image realism through advancements in GAN architectures, training, and data augmentation.

## References

1. Wang X, Guo H, Hu S, Chang MC, Lyu S (2022) GAN-generated FacesDetection: a survey and new perspectives. *Front Artif Intell Appl* 372:2533–2542. <https://doi.org/10.3233/FAIA230558>
2. Frowd CD et al (2005) Contemporary composite techniques: the impact of a forensically-relevant target delay. *Leg Criminol Psychol* 10(1):63–81. <https://doi.org/10.1348/135532504X15358>
3. Natural language processing (Almost) from scratch—academic torrents. Accessed 6 April 2024. [Online]. Available: <https://academictorrents.com/details/824fd119b03225610249c0ce6ceae778dc7e28d>
4. Devlin J, Chang MW, Lee K, Toutanova K (2018) BERT: pre-training of deep bidirectional transformers for language understanding, NAACL HLT 2019–2019 conference North American chapter association computer linguistics human language technology—proceeding conference, vol 1, pp 4171–4186. Accessed 6 April 2024. [Online]. Available: <https://arxiv.org/abs/1810.04805v2>
5. Ayanthi DMA, Munasinghe S (2022) Text-to-face generation with StyleGAN2, pp 49–64. <https://doi.org/10.5121/csit.2022.120805>
6. Chen X, Qing L, He X, Luo X, Xu Y (2019) FTGAN: a fully-trained generative adversarial networks for text to face generation. Accessed 6 April 2024. [Online]. Available: <https://arxiv.org/abs/1904.05729v1>
7. Kumar A, Mudgil A, Dodeja N, Vishwakarma DK (2021) Realistic face generation using a textual description. Proceeding—5th international conference computer methodology communication ICCMC 2021, pp 917–922. <https://doi.org/10.1109/ICCMC51019.2021.9418040>
8. Khan MZ et al (2021) A realistic image generation of face from text description using the fully trained generative adversarial networks. *IEEE Access* 9:1250–1260. <https://doi.org/10.1109/ACCESS.2020.3015656>
9. Nasir OR, Jha SK, Grover MS, Yu Y, Kumar A, Shah RR (2019) Text2FaceGAN: face generation from fine grained textual descriptions, Proceeding—2019 IEEE 5th international conference multimedia big data, BigMM, pp 58–67. <https://doi.org/10.1109/BigMM.2019.0042>
10. Jivarajani K (2023) Automatic synthesis of realistic human faces from text using GANs. *Int J Res Appl Sci Eng Technol* 11(5):7263–7271. <https://doi.org/10.22214/IJRASET.2023.53433>
11. Tao M, Tang H, Wu F, Jing X, Bao BK, Xu C (2020) DF-GAN: a simple and effective baseline for text-to-image synthesis, Proceedings IEEE computer social conference computer vision pattern recognition, vol 2022, pp 16494–16504. <https://doi.org/10.1109/CVPR52688.2022.01602>
12. Reimers N, Gurevych I (2019) Sentence-BERT: sentence embeddings using Siamese BERT-networks, EMNLP-IJCNLP 2019–2019 conference empirical methods natural language processing 9th international Jt. conference natural language processing proceeding conference, pp 3982–3992. <https://doi.org/10.18653/v1/d19-1410>
13. De Vries H, Strub F, Mary J, Larochelle H, Pietquin O, Courville A (2017) Modulating early visual processing by language, advance Neural information processing system, vol 2017, pp 6595–6605. Accessed 08 April 2024. [Online]. Available: <https://arxiv.org/abs/1707.00683v3>
14. Huang X, Belongie S (2017) Arbitrary style transfer in real-time with adaptive instance normalization. Proceeding IEEE international conference computer vision, vol 2017, pp 1510–1519. <https://doi.org/10.1109/ICCV.2017.167>
15. Karras T, Laine S, Aila T (2018) A style-based generator architecture for generative adversarial networks. *IEEE Trans Pattern Anal Mach Intell* 43(12):4217–4228. <https://doi.org/10.1109/TPAMI.2020.2970919>
16. Mescheder L, Geiger A, Nowozin S (2018) Which training methods for GANs do actually converge?. 35th international conference machine learning ICML 2018, vol 8, pp 5589–5626. Accessed 08 April 2024. [Online]. Available: <https://arxiv.org/abs/1801.04406v4>

17. Xu T et al (2017) AttnGAN: fine-grained text to image generation with attentional generative adversarial networks. Proceeding IEEE computer social conference computer vision pattern Recognition, pp 1316–1324. <https://doi.org/10.1109/CVPR.2018.00143>
18. Zhang H et al (2016) StackGAN: text to photo-realistic image synthesis with stacked generative adversarial networks, vol 2017, pp 5908–5916. Accessed 08 April 2024. [Online]. Available: <https://arxiv.org/abs/1612.03242v2>
19. CelebFaces attributes (CelebA) dataset. Accessed 08 April 2024. [Online]. Available: <https://www.kaggle.com/datasets/jessicali9530/celeba-dataset>

# Advanced Machine Learning Approach with Dynamic Analysis to Detect Malware in Cybersecurity Domain



**Shubhang Gupta, Shamim Khan, Tiansheng Yang,  
Rajkumar Singh Rathore, Aniket Das, and Nilamadhab Mishra**

**Abstract** This paper is about detecting malware using machine learning. As the fact that complex computer attacks are increasing rapidly, suggests that relying on old methods and techniques to identify them are insufficient. So to deal with this situation, machine learning has emerged as the most efficient approach in identification of malware. This paper starts with a brief introduction of different varieties of malware and how the danger is constantly evolving. Then, it explains basic principles and operations of ML models and how it learn from the existing data and mistakes. It discuss about several challenges associated with the use of ML technology in searching for malware. Finally, it talks about what could be done in the future to make it even better at detecting malware with the help of machine learning. This paper will also help other researchers, people who work in cyber security, and anyone else who wants to know more about using ML to look for malware.

**Keywords** Malware · Machine learning · Malware detection using ML · Challenges in malware detection

---

S. Gupta · S. Khan · A. Das  
Kalinga Institute of Industrial Technology, Bhubaneswar, India

T. Yang (✉)  
University of South Wales, Pontypridd, UK  
e-mail: [tiansheng.yang1@southwales.ac.uk](mailto:tiansheng.yang1@southwales.ac.uk)

R. S. Rathore  
Cardiff School of Technologies, Cardiff Metropolitan University, Llandaff Campus, Cardiff, UK  
e-mail: [rsrathore@cardiffmet.ac.uk](mailto:rsrathore@cardiffmet.ac.uk)

N. Mishra  
VIT Bhopal University, Sehore, Madhya Pradesh, India

## 1 Introduction

As we all know that malware is always changing rapidly, which makes it difficult for traditional cybersecurity methods to detect it. The old approaches are looking for similar signatures and patterns from the previous malware or from previous gathered data. However, signatures are irrelevant nowadays because signature-based approaches which search for certain patterns inside code, are no more sufficient as we all know that virus like polymorphic virus can make changes in there code during infest and become hard to detect. This is where machine learning helps a lot. It help by gathering old data and by recognizing patterns of the malware and help computer in differentiate the slight signs of malware even if they are trying to avoid it. Here, we'll discover how ML is revolutionizing cybersecurity, particularly in the field of malware discovery. First and foremost, we'll learn about several types of malware, such as viruses, Trojans, and ransomware, which will hold your data storage hostage. Recognizing these predators is half the warfare. Second, we'll learn about the fundamental concepts of machine learning like supervised, unsupervised and semi-supervised learning. These are the machines that enable computers to study from examples in the same way we do. Using machine learning is like providing our computers with a magnifying glass to spot concealed codes. Naturally, it is not that simple but with the help of machine learning it become much easier. Therefore, ML deals with dirty data and tough adversaries who seek to outwit our systems. Nonetheless, we will discover smart strategies to combat these problems, enhancing our defence further. So, let's get started with this ride of how machine learning deals with the malware.

## 2 Literature Review

Evolving malware presents an intense challenge to cybersecurity, as new threats emerge constantly. Conventional detection methods and techniques are unable to detect malware, which leads to several errors in the system. Machine learning emerges as a promising avenue to confront this challenge. It has the capability to learn and adapt from available data, empowering them to recognise similar patterns and behaviours characteristic of malware, notwithstanding its evolution [1].

### 2.1 Diverse World of Malware

Malware is coded software which use to damage any system. It is a malicious code which brings down system performance and some malware are used to make back door. Most of the malware comes from internet while downloading content from

the internet or sharing files. Different programming language can be used to create malware [2].

- Worms: It directly infects system file. It automatically creates file and time to time it spreads over the whole system.
- Trojan: It is a smart malware as it can change its appearance.
- Spyware: Malware designed to secretly monitor and collect information about users' activities, such as browsing habits or keystrokes.
- Adware: It enter into our system and shows random ads to generate revenue for the attackers.
- Spyware: It create a backdoor for the hacker to spy on the user's system and records all its activity.
- Rootkits: It hides itself in the system and takes user's access and controls over the system.

## 2.2 *Basics of Machine Learning*

Machine learning (ML) allows the computer to learn from the data which enables decision-making and understanding to a problem. It comprises of three major concepts: Supervised, unsupervised, and semi-supervised learning.

**Supervised Learning** [3]: It learn a function that maps and input to an output based on sample input–output pairs. For example, predicting a tweet or a product review.

**Unsupervised Learning** [4]: It processes the data without human interference. Here, some unsupervised learning tasks such as clustering, feature leaning, density estimation, and finding association rules.

**Semi-supervised Learning** [5]: It a combined form of both supervised and unsupervised learning. It provides a better outcome for prediction.

It enhances threat detection by analysing patterns in the cybersecurity datasets, which helps in empowering organisations to proactively identify and respond to evolving threats. The adaptability and scalability of machine learning outshine traditional signature-based methods, particularly in detecting subtle and polymorphic threats.

Despite the effectiveness of ML, it faces many challenges like handling noisy data and mitigating adversarial attacks.

Machine learning in cybersecurity holds promise with upcoming techniques like deep learning and explainable AI, which helps against cyber threats.

### 2.3 Machine Learning for Malware Detection

In cybersecurity, machine learning is used to detect malware. There are two main point regarding this: Static attributes and dynamic attributes [6].

1. Static Attributes: Malware has its own code and size which is detected by the ML algorithms by the use of techniques like n-gram analysis.
2. Dynamic Attributes: Any unusual activity from the malware will be detect by ML algorithms by using methods like sandboxing or runtime monitoring.

Algorithms taking both static and dynamic attributes detect ML-based malware. By now we can conclude that ML algorithms play a big role in malware detections which will help to stop dangerous attacks by the attackers.

### 2.4 Challenges in ML-Based Malware Detection

**Handling Messy Data:** Inconsistent and noisy data generated by ever-mutating polymorphic viruses [7], makes it difficult for ML algorithms to trace patterns efficiently due to humongous variations in behaviour and code.

**Adversarial Attacks:** Adversaries exploit faults in the model by designing different versions of malware successfully bypassing ML algorithms. They manipulate it to avoid being detected by ML-based systems [8].

To overcome these challenges, we have some strategies

**Robust Feature Engineering:** Rather than the code pattern, this extraction technique focuses on invariant characteristics of the malware which furthers detection accuracy. To generalise across polymorphic variants, we record high-level behavioural features like, system call sequences.

**Ensemble Learning:** These techniques can improve resilience and detection over adversarial attacks. The system can perform evasion attempts if we aggregate predictions from other models trained using various representations and features [9].

**Continuous Model Updating:** Mechanisms for continuous retraining and model updating to be implemented so that the ML system can adapt to rapidly evolving malwares. The system will eventually detect emerging variations of polymorphic virus more promptly if we incorporate new labelled data routinely.

**Adversarial Training:** It is a fact that to improve defence, ML models should be trained using adversarial examples by exposing it to malicious samples. Model will learn to mitigate and recognise evasion and will be robust in practical cases [10–12].

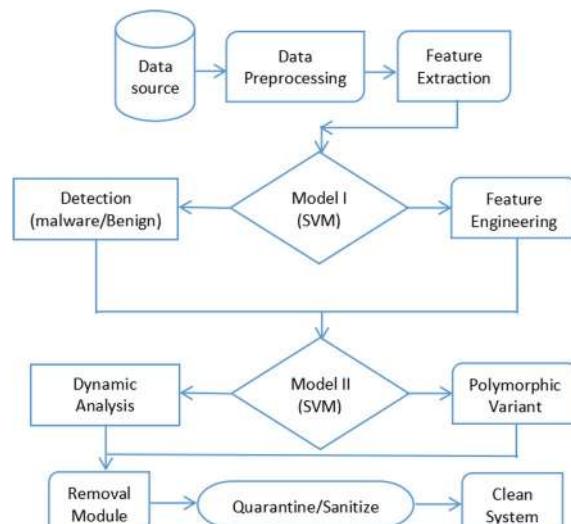
To combat the emerging cyber threats, the need of the hour is to address the above challenges. Polymorphic virus detection through ML will be effective and resilient if we implement the strategies discussed further [2, 13].

### 3 Proposed Methodology

The proposed methodology is shown in the figure.

1. Source: Files are received in form of various downloads, email attachments, etc. in our system through different sources.
  2. Pre-processing: Pre-processing tasks generally involve converting files to commonly used formats, unpacking archives or hashing which will prepare our file data for further analysis.
  3. Feature Extraction (Static Analysis): We classify files based on certain criteria and features like:
    - File metadata (size, creation date, etc.)
    - File headers and sections (PE headers for executables)
    - API calls used by the programme (for executables)
    - Code disassembly analysis (for complex malware) (Fig. 1)
  4. Model 1 (SVM): A Support Vector Machine (SVM) is used for initial classification. It's trained on a dataset of malware and benign files, learning to differentiate based on the extracted features.
  5. Detection (Malware/Benign): If the file is classified as malware, the system proceeds to the next stage.

**Fig. 1** Proposed model using machine learning for malware detection



6. Dynamic Analysis (if Malware): This stage involves running the file in a sand-boxed environment to observe its behaviour. It can involve monitoring the system calls made by the programme, network activity and file modifications.
7. Model 2 (LSTM): A Long Short-Term Memory (LSTM) network is used to analyse the dynamic behaviour sequence. LSTMs are adept at identifying patterns even in obfuscated or polymorphic code.
8. Polymorphic Variant: If the malware is a polymorphic variant, the model will further classify into the specific type [14].
9. Removal Module: System will take the below action based on the feature extraction results.
  - Quarantine: It ensures that the malware will not spread or affect other file systems as the infected files are to be transferred to some other secure place, isolated from remaining files avoiding further damage [15, 16].
  - Sanitise: In few cases, it is possible to preserve the original content of the file and simultaneously remove the malware by neutralizing the viral code. The legitimate data remains undamaged [17, 18].
10. Clean System: After the removal module quarantines or sanitises the infected files, a cleaning process is performed in the system to check if all minute traces of malicious code are removed. We must scan the entire system and take actions to check for remaining malware or other suspicious code, eliminating it finally.

## 4 Implementation and Discussion

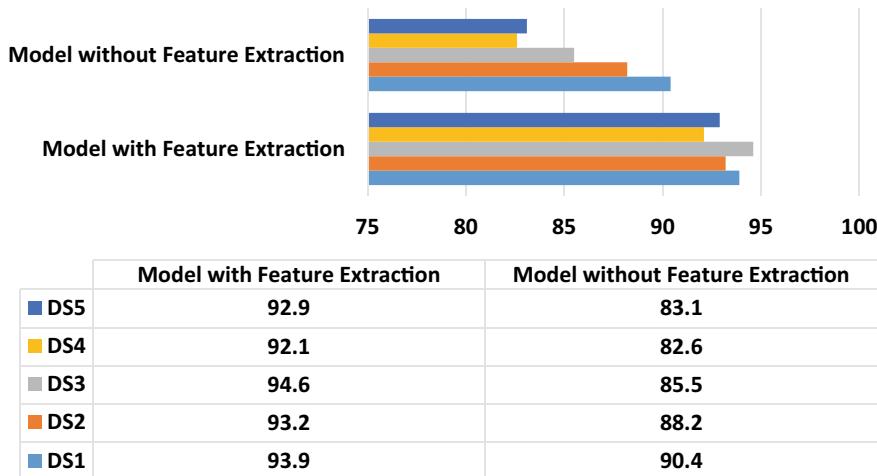
Outcome after implementing the proposed model is discussed as below.

To test the efficiency of malware detection, we compare it with other ML models. As evident from the table below, the best classifier is SVM. The recorded values using SVM for True Positive Rate (TPR), False Positive Rate (FPR), and Accuracy Rate (ACC) are 0.94, 0.91, and 0.95, respectively (Table 1).

When other records are compared with a particular model, its reliability is determined. As seen from Fig. 2, our proposed model is capable to generate a consistent f-score value with varying datasets of malware threat. An average f-score metric of 85.96% was noted with the proposed model.

**Table 1** Other ML malware detecting models performance metrics is shown below

Classifier	TPR	FPR	ACC
SVM	0.94	0.91	0.95
KNN	0.9	0.86	0.84
MLP	0.88	0.82	0.85
NB	0.81	0.78	0.8
DT	0.91	0.87	0.92



**Fig. 2** Proposed model f-score versus different malware datasets

## 5 Conclusion

By all counts and proven results, we can conclude that in today's world of cybersecurity detecting malware using machine learning become very necessary. However, it faces hurdles, especially with polymorphic viruses, which elude traditional detection methods. Despite these challenges, addressing issues like handling unstructured data and countering adversarial attacks can significantly bolster ML-based detection systems. Strategies such as robust feature engineering, ensemble learning, continuous model updating, and adversarial training enable ML systems to adapt to the dynamic nature of polymorphic viruses and other emerging threats. These tactics empower researchers and practitioners to strengthen ML algorithms in identifying and mitigating evolving malware variants. Although obstacles persist, the potential of ML in cybersecurity is immense. Ongoing refinement of detection techniques and proactive monitoring of emerging threats are crucial for enhancing cybersecurity defences. As our understanding and application of ML techniques advance, the future promises even more resilient cybersecurity solutions.

## References

1. Pandey VK, Prakash S, Gupta TK, Yang T, Singh A, Rathore RS (2024) A computational intelligence inspired framework for intrusion detection in WSN. In: 2024 International conference on decision aid sciences and applications (DASA). IEEE, pp 1–4
2. Mishra S, Chaudhury P, Tripathy HK, Sahoo KS, Jhanjhi NZ, Hassan Elnour AA, Abdelmaboud A (2024) Enhancing health care through medical cognitive virtual agents. Digital Health 10:20552076241256732

3. Tiwari PK, Prakash S, Tripathi A, Yang T, Rathore RS, Aggarwal M, Shukla NK (2025) A secure and robust machine learning model for intrusion detection in internet of vehicles. *IEEE Access*
4. Kashyap P, Pareek A, Mishra S, Khan Z, Garg R, Tripathy HK (2024) Sentiment polarity analysis of twitter data using machine learning models. In: International conference on innovative computing and communication. Springer Nature Singapore, Singapore, pp 623–635
5. Pradhan SR, Mishra S, Tripathy HK, Brahma B, Gobinath R, Sobti R (2024) Critical application feasibility of predictive learning in autonomous vehicles. In: International conference on innovative computing and communication. Springer Nature Singapore, Singapore, pp 371–383
6. Chakraborty S, Mishra S, Tripathy HK (2022) COVID-19 outbreak estimation approach using hybrid time series modelling. In: International conference on innovations in intelligent computing and communications. Springer International Publishing, Cham, pp 249–260
7. Sinha P, Prakash S, Jha SK, Rathore V, Yang T, Rathore RS, Singh A, Mishra R (2024) An efficient ML-based model for network intrusion detection system. In: 2024 International conference on decision aid sciences and applications (DASA). IEEE, pp 1–5
8. Singh AR, Kumar RS, Rathore RS, Pandian A, Alrayes FS, Allafi R, Ahmad N (2025) AI-enhanced smart grid framework for intrusion detection and mitigation in EV charging stations. *Alex Eng J* 115:603–621
9. Kumar V, Rathore RS (2018) Security issues with virtualization in cloud computing. In: 2018 international conference on advances in computing, communication control and networking (ICACCCN). IEEE, pp 487–491
10. Kumar S, Singh A, Benslimane A, Chithaluru P, Albahe MA, Rathore RS, Álvarez RM (2023) An optimized intelligent computational security model for interconnected blockchain-IoT system and cities. *Ad Hoc Netw* 151:103299
11. Patel AD, Jhaveri RH, Shah KA, Patel AD, Rathore RS, Paliwal M, Abhishek K, Thakker D (2024) Security trends in internet-of-things for ambient assistive living: a review. *Recent Adv Comput Sci Commun (Formerly: Recent Patents Comput Sci)* 17(7):18–46
12. Bhawana KS, Rathore RS, Mahmud M, Kaiwartya O, Lloret J (2022) BEST—blockchain-enabled secure and trusted public emergency services for smart cities environment. *Sensors* 22(15):5733
13. Sahoo S, Mishra S, Brahma B, Barsocchi P, Bhoi AK (2024) SSO-CCNN: a correlation-based optimized deep CNN for brain tumor classification using sampled PGGAN. *Int J Comput Intell Syst* 17(1):1–18
14. Kumar G, Rathore RS, Thakur K, Almadhor A, Biabani SAA, Chander S (2023) Dynamic routing approach for enhancing source location privacy in wireless sensor networks. *Wirel Network* 1–17
15. Mishra S, Jena L, Mishra N, Chang HT (2024) PD-DETECTOR: a sustainable and computationally intelligent mobile application model for Parkinson's disease severity assessment. *Heliyon* 10(14)
16. Pranjal P, Mallick S, Paul A, Mishra S, Bhardwaj I, Albuquerque VHCD (2024) Soil crops and nutrients forecasting using random forest model. In: AIP conference proceedings, vol 2919, no 1. AIP Publishing
17. Mishra S, Chakraborty S, Sahoo KS, Bilal M (2023) Cogni-Sec: a secure cognitive enabled distributed reinforcement learning model for medical cyber-physical system. *Int Things* 24:100978
18. Mishra S, Volety DR, Bohra N, Alfarhood S, Safran M (2023) A smart and sustainable framework for millet crop monitoring equipped with disease detection using enhanced predictive intelligence. *Alex Eng J* 83:298–306

# Enhanced Deepfake Detection Using Deep Learning on Large-Scale Video Data: A Fused ResNet50 and LSTM Approach



Naveen Ananda Kumar Joseph Annaiah and B. Omkar Lakshmi Jagan

**Abstract** In today's digital world, where what we see isn't always what it seems, the rise of Deepfake technology presents a big challenge to trust in videos. These manipulated videos, created using advanced AI, it helps in distinguishing between real and fake and makes it easier to spread false information. This study addresses the urgent need for reliable Deepfake detection methods using deep learning (DL) techniques within the Deepfake Detection Challenge (DFDC) dataset. Leveraging the vast and diverse DFDC dataset, this research develops robust detection algorithms capable of discerning between authentic and manipulated media content. Recently, researchers focused on Deepfakes detections, but accuracy is very low due to the content availability, rapid growth in digital content creation techniques. We proposed enhanced deep learning approach uses the fused ResNet50 and LSTM models, for more accurate detection of Deepfakes. When we evaluated the model over the DFDC dataset, it attains better F1-score, accuracy, and Logloss values compared to the existing models. The proposed model is intended to reduce the risks posed by Deepfake, including the spread of disinformation and the erosion of reputation and security. The study therefore helps in combating the negative impact of Deepfakes through the application of the DL models toward protecting trust, credibility and integrity in the digital space.

**Keywords** GAN · Deep learning · ResNet50 · LSTM · Deepfake · DFDC

---

N. A. K. J. Annaiah   
Tekinvaderz LLC, Florida, USA  
e-mail: [naveenjannaiah@gmail.com](mailto:naveenjannaiah@gmail.com)

B. O. L. Jagan  
Department of Electrical and Electronics Engineering, Vignan's Institute of Information Technology, Duvvada, Visakhapatnam, Andhra Pradesh, India

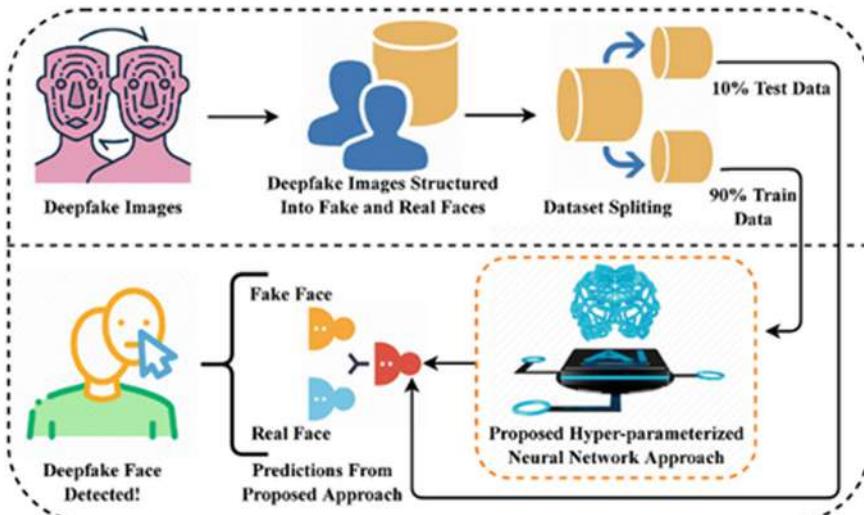
Department of Computer Science and Engineering, Vignan's Institute of Information Technology, Duvvada, Visakhapatnam, Andhra Pradesh, India

B. O. L. Jagan  
e-mail: [omkarjagan@yahoo.com](mailto:omkarjagan@yahoo.com)

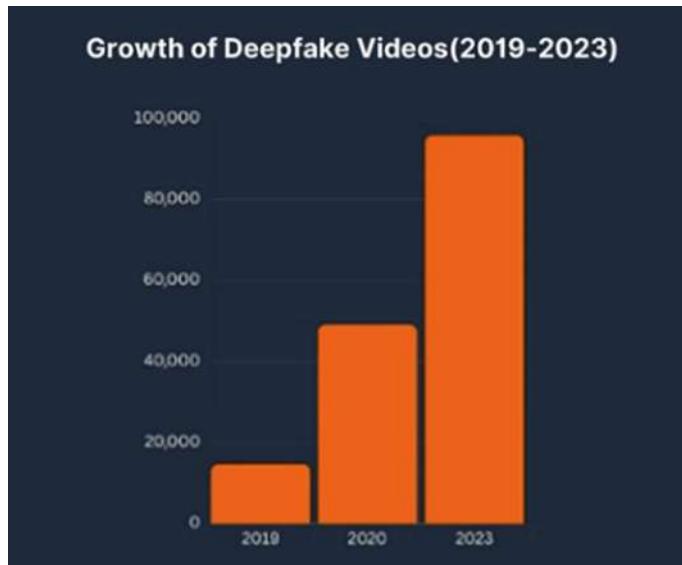
## 1 Introduction

Deepfake detection is a crucial step in the modern digital era where reality and illusion are difficult to distinguish. In this paper, we step into the study of Deepfake detection, utilizing DL algorithms to distinguish between videos which are real and those muted artificially. The study carried out on the large and diverse DFDC dataset, which also contains a rich collection of Deepfake videos. The dataset consists a huge number of Deepfake videos from different scenarios and wide range of appearances. Deepfake technology threatens not only individuals and organizations, but communities given how easily misinformation can spread due to the very serious implications of the technology such as defamation, damaged reputation or security compromise. This work can assist in mitigating these risks by developing strong detection algorithms that can detect such manipulated contents to some extent, hence ebbing those threats. Furthermore, the Deepfake technologies advancing every day and getting more sophisticated than ever before, so it becomes essential for equally advanced detection algorithms to keep pace. Apart from the Deepfake detection, this study also explores the digital ethics. Figure 1 depicts the process of Deepfake detection in real-world scenario.

Deepfakes became the huge potential threat to the integrity of digital media. Recent research tackles this problem by employing complex deep learning techniques to build strong Deepfake detectors. The DFDC dataset is an important resource for training and evaluating these models, where there has been a spike in the proliferation of Deepfake manipulation. This dataset provides labeled examples with varied quality, resolution, and manipulation videos. As shown in Fig. 2, the increasing



**Fig. 1** Deepfake detection process [1]



**Fig. 2** Number of cases recorded from 2019, 2020 and 2023 [2]

growth of Deepfake technology is a growing distress about its possible misuse for spreading the wrong information, influencing public opinions, and deceiving individuals, which creates a challenge for different regions such as journalism, politics, entertainment, and security.

The existing system for Deepfake detection typically depends on a combination of traditional image and video processing techniques, manual inspection, and forensic analysis. These methods often involve identifying inconsistencies such as unnatural facial movements, mismatches in lip-syncing, and artifacts introduced during the manipulation process. Additionally, some systems utilize metadata analysis and reverse engineering of the editing software to detect signs of tampering but all these approaches have several limitations. They can be time-consuming, manual labor, and may not be scalable to handle the volume and complexity of Deepfake content generated daily. Moreover, as Deepfake technology advances, traditional detection methods may become less effective at accurately identifying manipulated media. There are a lack of standardized datasets and benchmarks makes it challenging for comparing the performance of different detection mechanisms objectively. In the evaluation of the effectiveness of Deepfake detection systems different kind of metrics play an important role in determining the performance. These metrics serve as benchmarks for assessing the system's accuracy, robustness, and reliability.

The remaining paper is arranged in the following order:

Part 2 elaborates about the literature review, in which we discussed the findings of prior studies and the model they preferred to conduct this job and their creativity in the work.

Part 3 offers a quick report on deep learning.

Part 4 Discussion about the proposed model.

Part 5 has a detailed view of the experimental data analysis using our methodologies.

Part 6 discusses the problem description, conclusion, outcomes, and future scope.

## 2 Literature Review

Russ Howes et al. proposed the DFDC dataset [3] comprises 5K videos with two facial modification algorithms, accompanied by specific evaluation metrics and baseline performance from tested detection models. Wayne Wu et al. worked on the Deeper Forensics-1.0 [4], it has a very large dataset for Authenticity Verification in Real Face Imagery. This helps in establishing benchmarks and initial insights into the challenges of detecting Deepfakes, paving the way for further improvements in this field to improve the performance.

Lingzhi Li et al. developed Enhanced Face Forgery Detection using Face X-ray Technology [5], leveraging grayscale facial X-ray images to reveal if an input face image is a fusion of two distinct sources, achieving an impressive accuracy of 97.73% in identifying forgeries from prevalent face manipulation algorithms. Barsha Lamichhane et al. utilized both complete and sample datasets from the Deepfake Detection Challenge (DFDC) [6] to evaluate their model against pretrained models like VGG-19, Xception, and Inception-ResNet-v2. Their research explores diverse constraints including different resolutions while maintaining aspect ratios of 1:1 and 9:16, achieving an accuracy of 0.3756.

Sohail Ahmed Khan et al. proposed and developed a fused Convolutional Neural Network (CNN) approach utilizing VGG16, InceptionV3, and XceptionNet architectures and Resilience Through Fused CNN Predictions [7], to enhance Deepfake detection resilience, achieving a remarkable 96.50% accuracy on the dataset, surpassing other systems.

Nicol'o Bonettini et al. worked on Detecting Video Face Manipulations Using a CNN Ensemble Approach [8]. The researchers are working on detecting when faces in videos have been manipulated or altered using modern techniques. Authors used a wide variety of deep learning architectures for Deepfake detection and they also did the research by combining various CNN models. They used the EfficientNetB4 as a backbone network, incorporates two different approaches: Attention layers and Siamese training. Authors claimed that the ensemble of deep learning models works well in detecting manipulated phony faces. They've evaluated their approach against two datasets with more than 119,000 videos publicly available and the preliminary results are very encouraging. Authors claimed that their model attains the accuracy of 0.944.

Daniel Mas Montserrat et al. had developed an approach by combining CNN Ensemble Approach [9] with RNNs to classify facial features with temporal patterns from videos. It resulted in effective detection with competitive accuracy on the DFDC

dataset at 92.61%. The paper shows great efficacy of such an approach regarding the identification of Deepfake videos. They tested their approach on two datasets and attains 95% accuracy.

Young-Jin Heo et al. proposed a novel Vision Transformer model with distillation techniques [10] to enhance the detection of Deepfake videos effectively overcoming issues like overfitting and false negatives. Authors used the patch embedding as input and used the CNN for feature extraction. Authors claimed that their proposed Deepfake detection model outperforms the existing modern Deepfake detection techniques on the dataset with an AUC of 0.978 and an F1 score of 91.9 without ensemble techniques. Xiaodan Li et al. suggested the Partial Face Attack Challenge in Deepfake videos [11] and applied the Sharp Multiple Instance Learning approach (S-MIL) to propose how to handle such attacks. Researchers claimed that their model performed better than existing approaches on standard datasets. It includes direct mapping of instance embeddings into video-level predictions that holds it up to securing facial information and Deepfake detection.

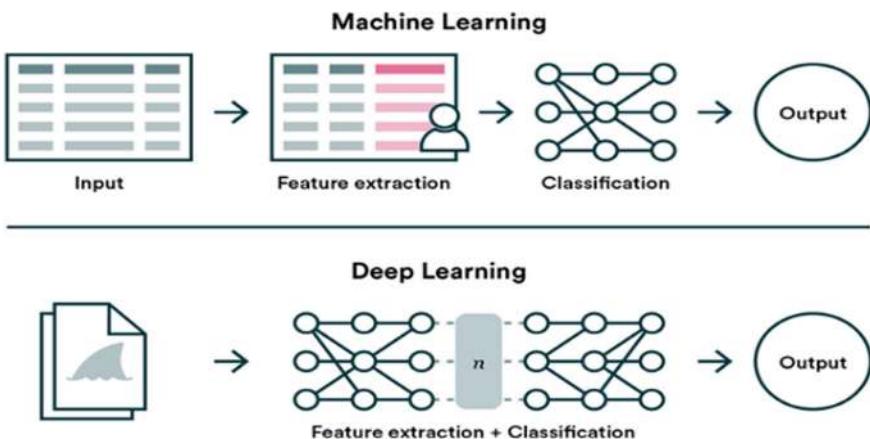
### 3 Deep Learning

DL is a subcategory of Machine Learning, designed toward direct understanding of varying complex patterns and representations directly from data. DL has exploded its popularity and achieved remarkable success and improvement on multiple benchmarks in various fields, including computer vision, NLP, speech recognition, and reinforcement learning. The major advantage of DL is its ability to automatically discover and learn features. This helps in making it particularly effective in handling large and complex datasets. It takes an inspiration from the structure and function of the human brain and it helps in enabling the machines to understand and learn from vastly different kinds of data. It helps in performing complex tasks with astounding performance. DL models have achieved highly developed performance in numerous amounts of tasks including image classification, game playing, object detection, and language translation.

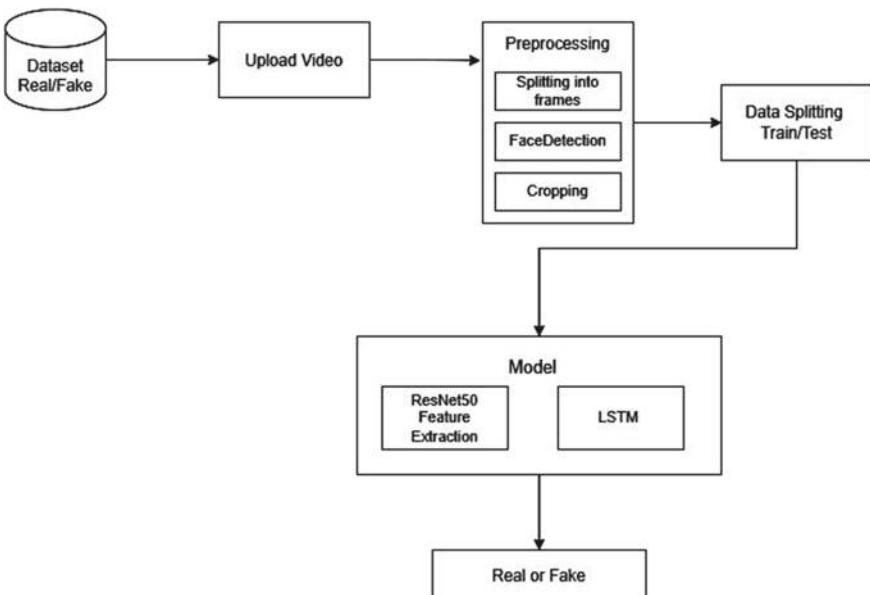
As shown in Fig. 3, deep learning can extract the feature from the images using various operations. Popular architectures in DL include CNN for image-related tasks, RNN for sequential data such as text or time-series data, and transformer models for NLP tasks. From image recognition and NLP to autonomous driving and healthcare.

### 4 Proposed Model

Deepfakes pose a significant challenge to digital media authenticity. This research explores the potential of deep learning to develop robust models capable of accurately distinguishing between genuine and manipulated media content. Figure 4 depicts the blueprint of the proposed Deepfake detection model.



**Fig. 3** ML and DL process



**Fig. 4** Proposed Deepfake detection model

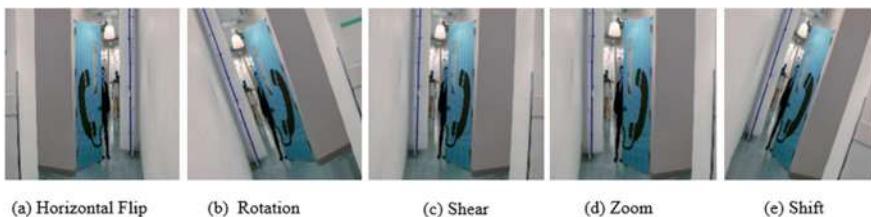
To detect the Deepfakes accurately, we need a large dataset of different kind of videos. We used the publicly available Deepfake Detection Challenge (DFDC) dataset. We applied various image pre-processing techniques, includes dividing videos into frames, detecting faces, and cropping the images before fed into the deep learning model.

## 4.1 Pre-processing

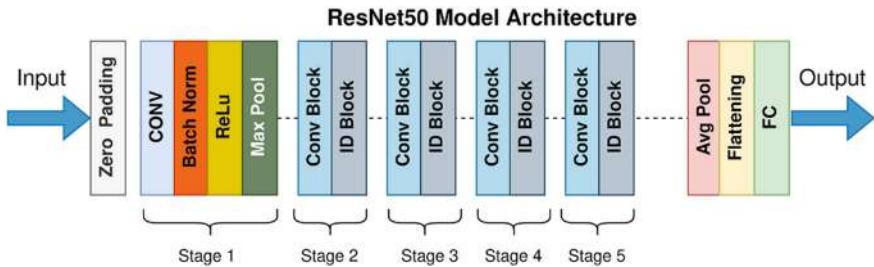
The pre-processing stage involves the extraction of frames that are uploaded from the dataset. The uploaded videos are divided into frames using the OpenCV library function. We utilize the OpenCV library to perform frame extraction, this process involves iterating through each frame of the input video and saving it as an image file. After all the frames are extracted then the process of face detection occurs by capturing face details for every frame. Detecting faces for video frames is done to extract relevant features for Deepfake detection. We employ face detection algorithms to locate and localize faces accurately. Once the detection of faces is completed, cropping is performed to focus on facial regions by eliminating irrelevant background information. Proper cropping ensures the model receives clean input data, enhancing its ability to distinguish between real and manipulated faces. At last, all the images are resized to the same size.

Figure 5 shows about flipping the image horizontally like looking at it in a mirror. The flipped image becomes an augmented version of the original image. And the image obtained after applying a rotation transformation to the image, rotating it by an angle of 20 degrees. The next image obtained after applying a shearing transformation to the image along a specified axis (typically horizontal or vertical). Next is Zoom means applying a scaling transformation to the image, either zooming in (enlarging) or zooming out (shrinking) the image. And in the above image, it displays the zoomed effect of the original image. The image depicts a translation transformation to the image, shifting its pixels horizontally.

Our model contains the integration of both CNN and RNN architectures. The CNN architecture is ResNet50 and the RNN architecture is Long Short-Term Memory (LSTM) is used to improve the performance of the architecture. The feature extraction is performed by Resnet50 architecture. ResNet50 is a variant of ResNets and consists of 50 layers. Residual Networks (ResNets) introduced a breakthrough in image classification tasks by reducing the vanishing gradient problem through skip connections. All the extracted features are used by LSTM model. LSTM is a type of recurrent neural network (RNN), specifically used for capturing dependencies over the long distribution of data. The fusion of ResNet50 and LSTM architectures involves integrating the ResNet50 layers with LSTM layers to enable end-to-end learning from sequential data. The ResNet50 component serves as a feature extractor, extracting



**Fig. 5** Augmented images



**Fig. 6** ResNet50 model [12]

hierarchical features from input data, while the LSTM component processes these features sequentially, capturing temporal dependencies and context information.

## 4.2 ResNet50

ResNet-50 is a CNN architecture that is part of the ResNet family. The name “ResNet50” originates from its structure as a residual network with 50 layers. ResNet50 belongs to the family of CNNs and has garnered widespread acclaim for its exceptional performance in image classification tasks. The fundamental building blocks of ResNet-50 are residual blocks, specifically the bottleneck residual block.

ResNet50 introduces skip connections to tackle the degradation problem. Figure 6 depicts the ResNet50 model architecture by utilizing the skip connections. The network can effectively learn the residual functions by referring to the inputs of each layer, instead of trying and learning the desired underlying mapping. The architecture of ResNet50 is structured around a series of residual blocks, each containing layers featuring batch normalization, multiple convolutional layers, rectified linear unit (ReLU) activations, and subsequently, shortcut connections to minimize the computational complexity.

## 4.3 Long Short-Term Memory (LSTM)

LSTM networks belong to the class of a RNN, designed in order to bridge the limitation posed by traditional RNNs in the finding the long-term dependency in sequential data. LSTM uses the ResNet50 architecture to overcome the vanishing gradient problem (VGP). By overcoming the VGP, the training model need not to change during the training process. LSTM uses various parameters like different learning rates, input and output biases, allows for flexibility while having control over the model’s behavior and training dynamics.

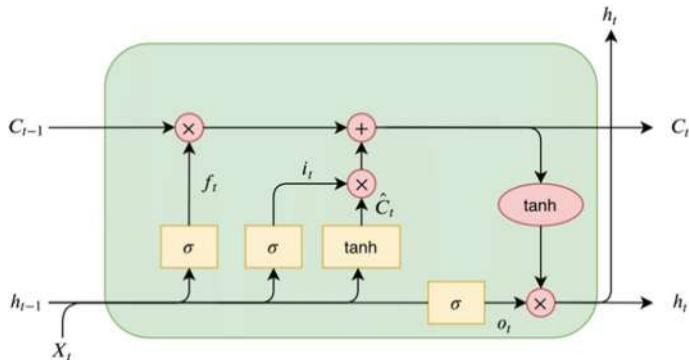


Fig. 7 LSTM model [13]

Figure 7 depicts the architecture of an LSTM model. In the LSTM networks, three types of gates are utilized: Which include forget gate, input gate, and output gate. These gates assist in integration for regulating the flow of data by allowing efficient memory management and learning. A standard LSTM network consists of several gates, which regulates the information flow at the cell state. The input gate is a kind of update to the cell state. Where the flow of information into the output is concerned, that is controlled by the output gate and the forget gate regularize the flow of information from the current cell to output cell state. We used LSTM-a variant of RNN architecture along with ResNet50 that has proven good to capture the temporal dependencies and patterns-when handling sequential data, especially in the context of sequence modeling tasks.

The entire training dataset need to be passed over the model. Passing of the model over the network called as an Epoch. Epochs represent the total number of forward passes and backward passes of the entire dataset during the training phase. If we define the number of epochs as 7 in the trained model, then the deep learning model will be trained for seven times over the training dataset. Increasing the number of epochs may lead to better performance with the model up to certain extent. The number of epochs can be as low as ten but up to 1000 or even higher depending on the complexity of the task and the convergence criteria.

## 5 Experimental Results and Analysis

### 5.1 Hardware Environment

The proposed model is executed on an Intel Core i7-12,700 CPU running at 2.10 GHz, provides quite strong computational power to complete the Deepfake detection tasks. The system configured with 32 GB ram and 2 GB Graphic card, allowed us to execute

computationally expensive operations, such as training deep learning models and working with big datasets in a timely and reliable way, while improving general performance and productivity in our computational workflows.

## 5.2 Software Environment

In this research, we used the well-known and most reliable Python libraries and frameworks not only for computer vision (CV) tasks but also to implement DL models. OpenCV (cv2) serves as an excellent basis for most of the image processing tasks like reading an image, pre-processing and augmentation of images. NumPy library is used to get the support for array manipulation and mathematical operations, which are most requisite operations for handling large, multi-dimensional data arrays. We used the Scikit-learn for pre-processing and splitting data into training and testing datasets.

TensorFlow is an open-source DL framework originally developed by Google and widely used in several ML and DL tasks. We also used the Keras, which reduces the creation and training time of artificial neural networks and provides an interface to work with TensorFlow.

## 5.3 Metrics

As the ResNet50 architecture employed in our model, residual blocks constitute the backbone, featuring convolutional layers alongside shortcut connections. Expanding upon the foundational ResNet design, ResNet50 incorporates refinements like pre-activation residual units. These units integrate batch normalization and ReLU activation prior to each convolutional operation, thereby enhancing the efficiency of training. The mathematical representation of a single residual block in ResNet can be represented as

$$H(X) = F(X) + X \quad (1)$$

As shown in Eq. (1),  $F(X)$  represents residual mapping,  $X$  represents input of the block,  $H(X)$  represents output of the block. While the equations provided describe the mathematical operations of a generic LSTM unit, the TensorFlow library handles the implementation of these operations which internally performs the mathematical operations necessary for the LSTM to function.

$$\text{Input Gate: } i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (2)$$

$$\text{Forget Gate: } f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (3)$$

$$\text{Output Gate: } o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (4)$$

$$\text{Candidate Cell State: } \tilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (5)$$

$$\text{New Cell State: } c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \quad (6)$$

$$\text{Hidden State Output: } h_t = o_t \cdot \tanh(c_t) \quad (7)$$

As shown in Eqs. (2–7) described above,  $\sigma$  represents the sigmoid activation function,  $\tanh$  denotes the hyperbolic tangent activation function.  $x_t$  represents the input at time step  $t$ ,  $h_t$  is the hidden state at time step  $t$ ,  $c_t$  is the cell state at time step  $t$ ,  $i_t, f_t, o_t, \tilde{c}_t$  and, respectively, represent the input gate, forget gate, output gate, and candidate cell state at time step  $t$ . The  $W$  matrices and  $b$  vectors are the weight matrices and bias vectors, respectively, which are the learnable parameters of the LSTM layer.

## 5.4 Results

The proposed fused ResNet50 and LSTM deep learning model demonstrated superior performance compared to state-of-the-art models across multiple evaluation metrics. Notably, our model achieved a significantly higher accuracy, F1-score, and lower logloss on the DFDC dataset. These results underscore the effectiveness of combining the powerful feature extraction capabilities of ResNet50 with the sequential learning advantages of LSTM for the given task. As shown in Table 1, the table gives a comparative study of accuracy, F1-score and Logloss values of different existing models and the proposed model.

The proposed fused ResNet50 and LSTM approach can identify the Deepfakes efficiently. Due to the advancements in technology the quality of Deepfakes are increasing rapidly, so more robust systems are need to be developed.

**Table 1** Accuracy, F1-score, and Logloss values of different models

Model	Accuracy (%)	F1-score (%)	Logloss
EfficientNet [8]	87	–	0.4658
VSS16 + InceptionV3 + XceptionNet [7]	96	–	0.11140
EfficientNet [10]	97.8	91.9	–
VGG-19 [6]	37	26.8	–
Face X-ray [5]	80.92	–	–
Our proposed model	93.4	90.6	0.0115

## 6 Conclusion

Despite the progress made in Deepfake detection, we have created a model architecture to find whether a video is real or fake. We have used ResNet50 and LSTM models to improve the efficiency of the project. By combining these models, we have achieved notable success in differentiating between genuine and manipulated videos. Moving forward, there are several avenues for further research and improvement in Deepfake detection using ResNet50 and LSTM. Firstly, exploring ensemble approaches that combine multiple detection models could enhance robustness and generalization. Secondly, investigating the integration of other modalities, such as audio and text, could provide complementary information for more accurate detection. Additionally, developing techniques to mitigate the impact of adversarial attacks on detection models is crucial. Furthermore, addressing ethical considerations and privacy concerns associated with Deepfake detection more robust deep learning models and algorithms are need to be improved.

## References

1. Raza A, Munir K, Almutairi M (2022) A novel deep learning approach for deepfake image detection. *Appl Sci* 12(19):9820. <https://doi.org/10.3390/APP12199820>
2. Lewis A, Vu P, Duch RM, Chowdhury A (2023) Deepfake detection with and without content warnings. *R Soc Open Sci* 10(11). <https://doi.org/10.1098/RSOS.231214>
3. Dolhansky B et al (2020) The DeepFake detection challenge (DFDC) dataset. Accessed 04 April 2024. [Online]. Available: <https://arxiv.org/abs/2006.07397v4>
4. Jiang L, Li R, Wu W, Qian C, Loy CC (2020) DeeperForensics-1.0: a large-scale dataset for real-world face forgery detection. *Proceedings IEEE computer social conference computer vision pattern recognition*, pp 2886–2895. <https://doi.org/10.1109/CVPR42600.2020.00296>
5. Li L et al (2019) Face X-ray for more general face forgery detection. *Proceeding IEEE computer social conference computer vision pattern recognition*, pp 5000–5009. <https://doi.org/10.1109/CVPR42600.2020.00505>
6. Lamichhane B, Thapa K, Yang SH (2022) Detection of image level forgery with various constraints using DFDC full and sample datasets. *Sensors* 22(23):9121. <https://doi.org/10.3390/S22239121>
7. Khan SA, Artusi A, Dai H (2021) Adversarially robust deepfake media detection using fused convolutional neural network predictions. Accessed 04 April 2024. [Online]. Available: <https://arxiv.org/abs/2102.05950v1>
8. Bonettini N, Bondi L, Cannas ED, Bestagini P, Mandelli S, Tubaro S (2020) Video face manipulation detection through ensemble of CNNs. *Proceeding—international conference pattern recognition*, pp 5012–5019. <https://doi.org/10.1109/ICPR48806.2021.9412711>
9. D. M. Montserrat et al (2020) Deepfakes detection with automatic face weighting. *IEEE computer social conference computer vision pattern recognition work*, vol 2020, pp 2851–2859. <https://doi.org/10.1109/CVPRW50498.2020.00342>
10. Heo Y-J, Choi Y-J, Lee Y-W, Kim B-G (2021) Deepfake detection scheme based on vision transformer and distillation. Accessed 04 April 2024. [Online]. Available: <https://arxiv.org/abs/2104.01353v1>
11. Li X et al (2020) Sharp multiple instance learning for DeepFake video detection. *MM 2020—Proceeding 28th ACM international conference multimedia*, pp 1864–1872. <https://doi.org/10.1145/3394171.3414034>

12. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. Proceeding IEEE computer social conference computer vision pattern recognition, vol 2016, pp 770–778. <https://doi.org/10.1109/CVPR.2016.90>
13. Long Short-Term Memory (LSTM). Long Short-Term Memory (LSTM) is a type... | by Saba Hesaraki | Medium. Accessed 04 April 2024. [Online]. Available: <https://medium.com/@saba99/long-short-term-memory-lstm-fffc5eaebfdc>

# Advancements in Digital Pathology: A Comprehensive Survey of Predictive Models for Cancer Diagnosis



K. Amuthachenthiru and M. Kaliappan

**Abstract** In the domain of cancer prediction using pathology of computation. Starting with automating routine diagnostic procedures and algorithms for the analysis of histopathology images has been notable been noteworthy. Artificial intelligence, or AI, has advanced, with applications ranging from finding prognostic and predictive indicators derived from tissue structure to automating routine diagnostic tasks. A number of obstacles, including ethical, operational, technical, cultural, financial, and regulatory elements and dimensions, stand in the way of the integration of computational pathology into clinical settings, despite its enormous promise. This survey outlines the present state of translational medicine from the perspective of pathologists investigation, assessing clinical application, and tackling common issues impeding broad clinical use. Modern approaches to help with the use of computational pathology methods are also covered in the survey. This paper examines how digital technologies are changing pathology, with a particular emphasis on cancer detection prediction models. With the integration of machine learning (ML) and artificial intelligence (AI) tools, pathology has entered a new era of greater efficiency and accuracy in cancer diagnosis. Digital pathology slide digitization is made possible by the use of whole-slide imaging (WSI).

**Keywords** Artificial intelligence · Machine learning · Support vector machines · Whole-slide image · Convolutional neural networks · Immunohistochemistry

---

K. Amuthachenthiru (✉)

Department of Artificial Intelligence and Machine Learning, R.M.D. Engineering College,  
Thiruvallur, India

e-mail: [amuthachenthiru@ritrjpm.ac.in](mailto:amuthachenthiru@ritrjpm.ac.in)

M. Kaliappan

Department of Artificial Intelligence and Data Science, Ramco Institute of Technology,  
Rajapalayam, Tamil Nadu, India

e-mail: [kaliappan@ritrjpm.ac.in](mailto:kaliappan@ritrjpm.ac.in)

## 1 Introduction

Even though the death rate from cancer has decreased by 2.2% since 2016, it is still the second leading cause of death in the US. An estimated 1.9 million Americans will be diagnosed with cancer in 2021 [1]. Cancer remains a serious global health concern advanced diagnostic tools are needed to address this issue and improve the effectiveness of treatment and early detection. Because computational methods have the potential to revolutionize cancer prediction, they have garnered a lot of attention lately, especially in pathology diagnostic report systems. This introduction explores the area of cancer prediction from a pathology perspective, highlighting the influence of state-of-the-art technologies on the development of diagnostic techniques.

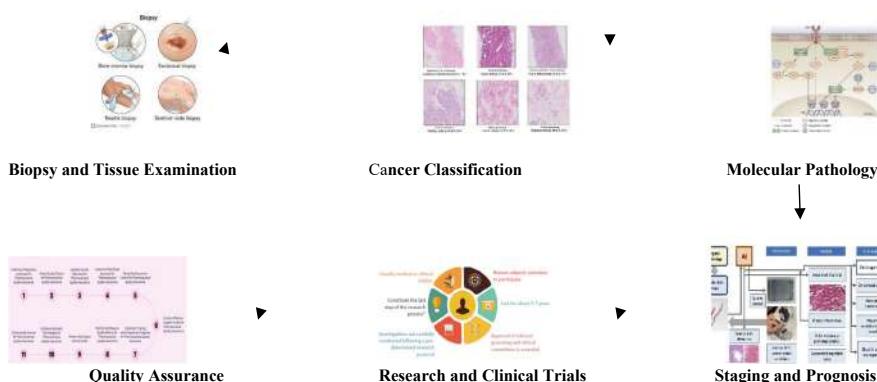
### 1.1 Methodology

The area of medicine known as “cancer pathology” deals with the investigation of diseases with a particular emphasis on the nature and causes of cancer. In order to detect the existence of cancer, identify its type, grade, and stage, and provide important information for treatment planning, pathologists analyze tissue and cell samples taken during biopsies or surgical operations.

An outline of significant facets of cancer pathology and the significance of precise diagnostic reports is provided (Fig. 1).

#### Biopsy and Tissue Examination

**Biopsy:** A biopsy entails taking a small sample of tissue or cells from the suspicious area and having it examined under a microscope. **Histopathology:** To find anomalies suggestive of cancer, pathologists examine the architecture of the tissue, the morphology of the cells, and other characteristics.



**Fig. 1** Comprehensive landscape: Biopsy to prognosis in oncology

## Cancer Classification

Pathologists play a crucial role in cancer diagnosis and treatment planning because they categorize malignancies according to their histological type and grade. Whereas grade evaluates the degree of abnormalities or aggressiveness of the cancer cells, histological classification identifies the precise tissue or cell type from which the cancer originates. Oncologists need this information in order to create personalized treatment plans that are effective for each individual malignancy.

### Histological Classification

**Adenocarcinoma:** Found frequently in organs such as the breast, prostate, and colon, adenocarcinomas originate from glandular tissues. **Squamous Cell Carcinoma:** These malignancies, which are derived from squamous epithelium cells, are frequently detected in the skin, lungs, and other organs.

**Leukemia:** A type of cancer that affects the bone marrow and blood that is characterized by the rapid production of abnormal white blood cells. **Lymphoma:** Cancers that start in the lymphatic system include both non-Hodgkin's and Hodgkin's lymphomas. **Melanoma:** Melanocytes are cells that create pigment and are typically found in the skin.

### Grading

**Well-Differentiated (Low Grade):** Malignant cells closely resemble typical cells and have a tendency to proliferate and disseminate at a less rapid pace. **Moderately Differentiated (Intermediate Grade):** Cells show some abnormal features but are not as aggressive as poorly differentiated cells.

**Poorly Differentiated (High Grade):** Cancer cells appear highly abnormal and tend to grow and spread rapidly.

### Molecular Pathology

Molecular pathology, which explores the genetic and molecular changes that underpin cancer development, is a critical advancement in cancer diagnoses and treatment. Molecular pathology examines the particular mutations, gene expressions, and molecular markers present in cancer cells in addition to classical histology. This thorough investigation not only improves the way cancers are classified, but it also offers important new information on how certain tumors behave. Oncologists can more precisely customize treatment plans by identifying unique molecular profiles. For example, the identification of particular mutations may direct the use of targeted medicines, which specifically target the faulty processes causing cancer to spread. Molecular pathology's inclusion in cancer treatment thus represents a paradigm change in favor of personalized medicine.

### Staging and Prognosis

A crucial component of oncology that pathologists manage is cancer staging, a methodical procedure that assesses how far along the disease has gone throughout the body. Examining the main tumor's size, extent of invasion into surrounding tissues,

involvement of lymph nodes, and possible metastases in distant organs are all part of this thorough evaluation. In addition to helping to classify the cancer's severity, staging information is an essential prognostic factor that offers important insights into how the disease is expected to progress. Staging helps physicians create treatment regimens that are suited to the unique features of the cancer by guiding choices about radiation therapy, chemotherapy, surgery, or a mix of these modalities. Furthermore, the cancer's stage is crucial in determining the overall prognosis, benefiting both.

## ***1.2 Treatment Planning***

Precise pathology reports are essential for customized cancer treatment planning because they give oncologists the knowledge they need to choose the best therapeutic approaches. The pathology results include information regarding the histological type, grade, molecular features, and stage of the disease. This information is useful in customizing treatment plans to the particular characteristics of each patient's cancer. For example, the choice of targeted therapies may be influenced by the discovery of particular mutations or molecular markers, and the stage of the cancer affects the choice of surgery, radiation therapy, or systemic treatments like immunotherapy and chemotherapy. With the use of these insights, oncologists can create complete, individualized treatment programs that maximize the trade-off between reducing possible adverse effects and increasing therapeutic efficacy.

## ***1.3 Monitoring Treatment Response***

By assessing tissue samples taken either during or after cancer therapy, pathologists are essential in tracking the effectiveness of treatment. These evaluations play a crucial role in determining how well the selected treatment modalities—such as immunotherapy, chemotherapy, radiation therapy, and surgery—work. Pathologists can offer important insights into whether the treatment is accomplishing its intended aims, such as reducing tumor size, removing cancer cells, or stopping further advancement, by examining alterations in the cancer cells and surrounding tissues. Oncologists need the data from these evaluations in order to make well-informed judgements regarding changing or modifying the current treatment strategy. The pathology results can validate the selected technique and direct decisions regarding treatment intensity modification or continuation if they show a positive response.

## ***1.4 Research and Clinical Trials***

Pathology is essential to the advancement of cancer research because it provides important insights into the complex biology of various cancer forms. Pathologists identify important molecular and genetic traits that drive the emergence and spread of malignancies by analyzing tissue samples. These findings contribute to our understanding of the disease and lay the foundation for the design and justification of clinical trials aimed at developing novel cancer treatments. Pathological data are essential in identifying possible therapy targets, biomarkers, and prognostic indications. Pathology plays a crucial role in patient stratification within clinical trials, guaranteeing that participants are suitably chosen according to the genetic and histological characteristics of their tumors. Pathological evaluations both during and following clinical trials offer crucial therapeutic input.

## ***1.5 Patient Communication***

For healthcare providers and patients to communicate effectively, pathology reports must be accurate. Pathology reports help patients comprehend their diagnosis, prognosis, and available treatments.

### **Quality Assurance**

In pathology labs, stringent quality control protocols are required to ensure the accuracy and consistency of diagnostic findings.

Continuous advancements in technology and ongoing training for pathologists contribute to improved diagnostic accuracy.

This survey paper aims to navigate the current landscape of cancer prediction within pathology diagnostic report systems. Through a thorough synthesis of extant literature and research, our aim is to furnish a comprehensive synopsis of the function that predictive models perform in enhancing diagnostic precision. Through an exploration of methodologies, challenges, and breakthroughs, our survey aims to contribute to the understanding of how predictive models can enhance cancer diagnosis in pathology.

## 2 Background

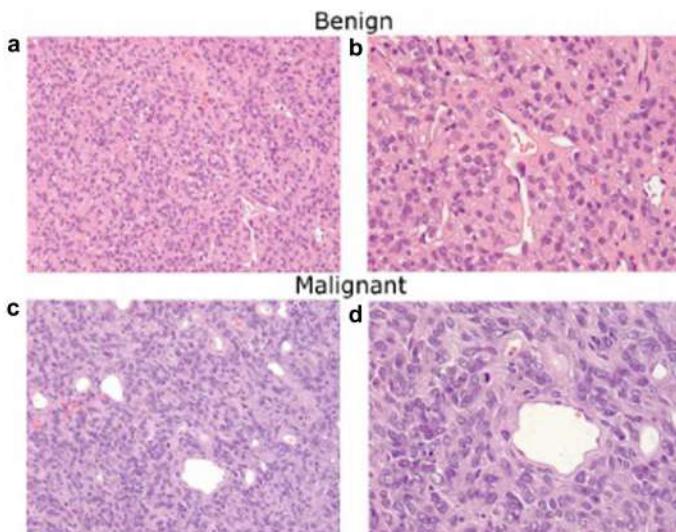
### 2.1 Methods of Cancer Diagnosis in Pathology

#### Histopathology

Histopathologists manage the cells and tissues removed from questionable “protuberances and irregularities” to provide a cancer diagnostic service. They identify the type of abnormality and, if they are malignant, provide the physician with information about the type, grade, and, in some situations, response to particular cancer therapies. Innovative imaging techniques can now obtain biopsy material from previously unreachable areas like the pancreas or retroperitoneum. A nighttime microscope examination of processed tissue is typical. The specimen can be quickly examined under certain conditions and with the use of specialized techniques. Pathologists are developing cutting-edge methods to analyze the genetic material in tissues or tumors, like polymerase chain reaction (PCR) and fluorescence in-situ hybridization (FISH), due to the rapid advancements in molecular pathology. These methods are essential to the treatment of many types of cancer (Fig. 2).

#### Immunohistochemistry

Pathologists use immunohistochemistry (IHC) as a laboratory technique to look for signs of disease in tissue samples. Lab tests are used by pathologists to diagnose diseases. A biopsy is a procedure in which a medical professional removes tissue and sends it to a lab for analysis. An IHC is one method used to examine the sample



**Fig. 2** Pathological dichotomy

after it gets to the lab. Using specific markers and antibodies to “label” the constituents of a tissue sample helps pathologists identify samples more easily. IHC is the most common type of immunostaining.

**Diagnostic Purpose:** Healthcare professionals can diagnose diseases, including cancer, thanks to IHC. It helps identify which kind of cancer, such as sarcoma, melanoma, or carcinoma, is present. It also helps to determine the causes of metastatic cancer.

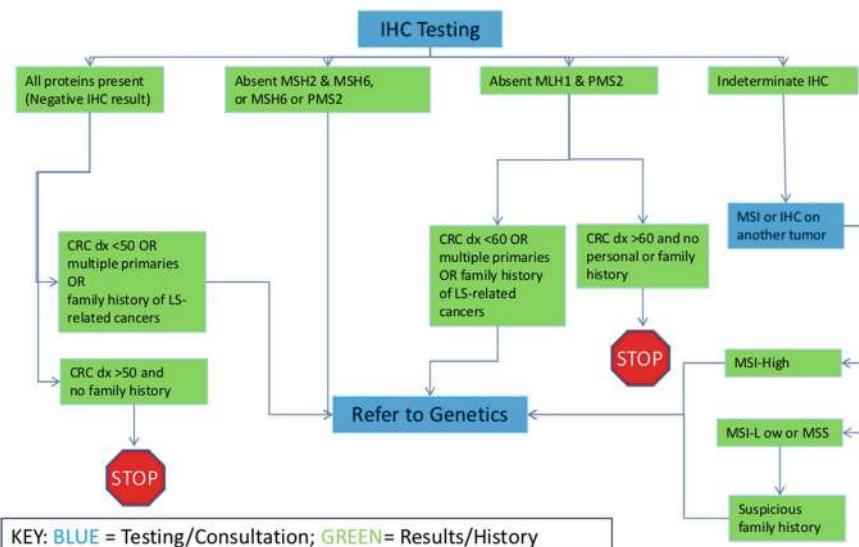
**Prognostic Assessment:** IHC is crucial for determining the aggressiveness or risk of malignancy. It provides important information for choosing the best course of treatment and helps with cancer staging and grading.

**Prediction of Treatment Response:** Tumor tissue features can be identified by IHC, providing information on potential therapy responses for cancer. For example, it helps pathologists identify breast and prostate cancers that are more prone to spread when particular hormones, such as testosterone and estrogen, are present. This information helps guide the administration of hormone-blocking treatments.

**Monitoring Treatment Progress:** Healthcare professionals can use IHC as a method to track how well a treatment is curing a patient’s illness (Fig. 3).

### Molecular Pathology

Molecular pathology explores cellular alterations, primarily concentrating on the therapeutic aspects of cancer treatment. The Molecular Pathology Unit at The Mount Sinai Hospital closely studies alterations in cancer cells to better assist patients. Lung

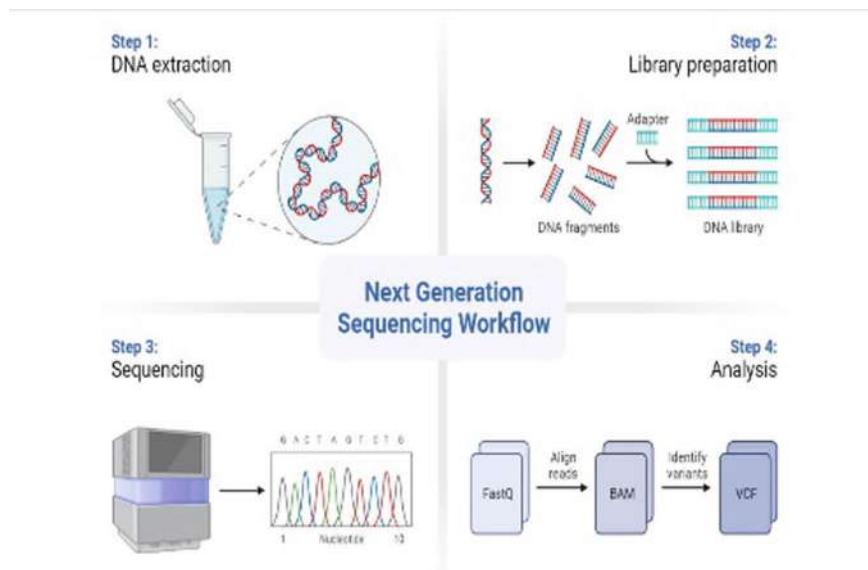


**Fig. 3** IHC testing process

cancer and breast cancer are currently the two most common scenarios for molecular pathology investigations, but assessments are expanding to encompass a number of other conditions, such as HPV in head and neck cancer. For instance, our molecular pathologists can identify which patients would benefit more from chemotherapy and which patients with a particular type of lung cancer histology might respond better to oral medication by examination of a tissue sample from a lung tumor. This method maximizes time efficiency while avoiding pointless treatments. The procedure entails analyzing modifications in cancer cells at various levels, from DNA to proteins, and predicting the cells' reactions to specific drugs.

### Next-Generation Sequencing (NGS)

Next-generation sequencing (NGS), deep sequencing, and massively parallel sequencing (MPS) are terms that are related to a DNA sequencing technology that has transformed genomic research. One day can be utilized to sequence the entire human genome using NGS. On the other hand, the final draft of the human genome was delivered more than ten years ago using the Sanger sequencing technology. In genome research, NGS has largely replaced traditional Sanger sequencing; however, it has not yet been adopted into standard clinical practice. Reviewing the possible uses of NGS in pediatrics is the goal of this article (Fig. 4).



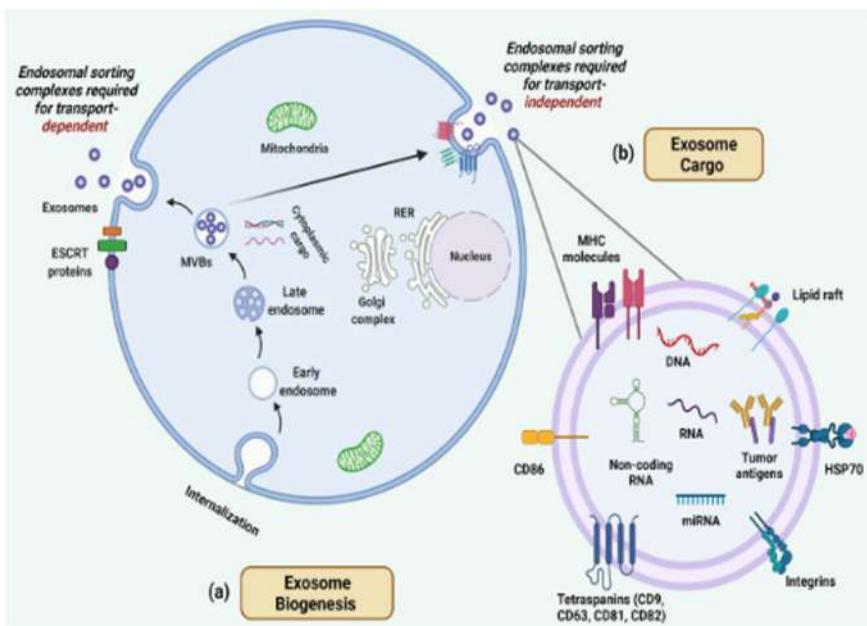
**Fig. 4** Workflow of next-generation sequencing

## Cytology

In cytology, a single cell type that is frequently found in fluid samples is analyzed. Its primary purpose is the identification or assessment of cancer. Additionally, it serves in screening for fatal irregularities, conducting pap smears, diagnosing infectious agents, and in various screening and diagnostic capacities.

### Liquid Biopsy

A fluid biopsy refers to a blood examination capable of identifying cancer cells or DNA in circulation, termed “circulating tumor DNA” or “ctDNA.” Similar to normal cells, cancer cells undergo a life cycle of death and replacement. As these deceased cells disintegrate, they are discharged from the tumor and enter the bloodstream. The liquid biopsy identifies the minute DNA fragments in the bloodstream originating from these cancer cells (Fig. 5).



**Fig. 5** Process of liquid biopsy

### 3 Challenges and Limitations of Traditional Diagnostic Approaches

S. No.	Approach	Challenges	Limitations
1	Biopsy-based diagnoses	Invasive procedures, such as biopsies, are the primary means of confirming cancer diagnoses	Biopsies may pose risks, are uncomfortable for patients, and might not capture the heterogeneity of tumors
2	Limited sensitivity for early detection	Traditional diagnostic methods may lack the sensitivity required to detect cancer at its early stages	Late-stage diagnoses can significantly impact treatment outcomes and reduce the effectiveness of therapeutic interventions
3	Subjectivity in pathological evaluation	Pathological evaluation, often subjective, can lead to interobserver variability	Inconsistencies in interpretation may result in misdiagnoses or delayed treatment initiation
4	Inadequate monitoring of treatment response	Monitoring treatment response traditionally relies on periodic imaging and clinical assessments	This approach may not provide real-time information, hindering prompt adjustments to treatment plans
5	Resource-intensive imaging techniques	Advanced imaging techniques, while effective, can be resource-intensive and costly	Limited accessibility to these techniques may impact their widespread adoption, particularly in resource-constrained settings
6	Challenges in liquid biopsy adoption	Liquid biopsy as a non-invasive alternative faces challenges in standardization and widespread adoption	Issues related to sensitivity and specificity may hinder the reliability of liquid biopsy results
7	Integration of molecular markers	Incorporating molecular markers into traditional diagnostic approaches requires validation and standardization	Despite promising advancements, the integration of molecular markers may not be universally adopted, affecting diagnostic accuracy

### 4 Literature Review

Reference [2] Deep learning (DL) is being incorporated into oncology to tackle the problems caused by the growing number of complex molecular biomarkers, which frequently lead to higher costs and longer times for making decisions in

routine clinical practice. Deep learning (DL) is an artificial intelligence technique that shows potential for directly extracting useful information from routine cancer histology images, making better use of the tumor tissue already present. While advanced DL approaches delve into molecular feature inference, survival prediction, and end-to-end therapy response prediction, potentially influencing clinical decision-making, basic image analysis tasks, like tumor detection and grading, aim to automate pathology workflows. Since more and more prognostic and predictive biomarkers are influencing oncology decision-making, which is currently complex and nonlinear, accurate decision-making systems are essential. These systems have the potential to enhance and accelerate decision-making processes if they undergo a thorough external validation procedure in clinical settings. Furthermore, the prevalence of molecular subpopulations in solid tumors that are the focus of novel therapeutic agents highlights how crucial DL is to the design of clinical trials in order to efficiently screen participants. However, the accuracy of DL predictions must be thoroughly validated to ensure their reliability and effectiveness in guiding personalized cancer treatments.

Reference [3] Systems Pathology signifies a procedural transformation in the execution of traditional diagnostic pathology, departing from the conventional use of deparaffinized tissue sections. The described technological progress serves as instances of methods through which the phenotypic attributes present in diagnostic specimens can be consistently extracted and applied in the clinical realm for regular patient care. We are of the opinion that integrating a systems-oriented strategy via operational histology will establish a structure for progressing personalized medicine and designing targeted therapies more comprehensively.

Reference [4] Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning, gastric cancer ranks fifth globally in cancer incidence and third in cancer-related deaths, with the highest rates in East Asian populations, particularly in China. A shortage of anatomical pathologists poses a challenge for timely and accurate diagnosis. Digital slides, especially whole-slide images (WSIs), and artificial intelligence (AI) assistance systems, particularly those using deep learning, show promise in addressing these challenges. The study details the deployment of an AI support system at the Chinese PLA General Hospital that uses a CNN to detect cancer at the pixel level in gastric pathology. The system demonstrates high sensitivity (0.996) and average specificity (0.806) on a daily dataset, indicating potential improvements in diagnostic accuracy and reduction of misdiagnoses when used by pathologists. A multicenter test across different hospitals confirms the system's robustness, highlighting its potential for widespread clinical application.

Reference [5] In computational pathology, histopathology images are analyzed and understood through the use of deep learning techniques and algorithms. Technological developments in AI have caused an explosion of innovative discoveries in computational pathology, ranging from the potential mechanization of standard diagnostic procedures to the identification of novel tissue-based prognostic as well as predictive biomarkers. Despite the enormous potential of computational pathology, its integration into clinical settings has proven challenging due to a number of issues, that includes operational, technical, regulatory, ethical, financial, and cultural ones.

We focus on pathologists' perspectives of computational pathology in this context, outlining its current state in translational research, evaluating its clinical utility, and addressing the common obstacles to clinical adoption and application. We wrap off by outlining current methods to advance these approaches.

Reference [6] Accurate diagnosis and prognosis prediction of gastric cancer using deep learning on digital pathological images: A retrospective multicenter study, the study addresses the critical need for accurate prognostic tools in gastric cancer (GC) to guide clinicians in determining the necessity for adjuvant treatments. Recognizing the limitations of manual histological analysis, the researchers employ deep learning, specifically convolutional neural network (CNN) technology, to develop two models: GastroMIL for the diagnosis of GC from pathological images and MIL-GC for predicting patient outcomes. The models outperform junior pathologists and align with expert pathologists, exhibiting high diagnostic accuracy and prognostic performance after being trained on a sizable dataset from multiple centers. The study introduces a user-friendly online platform for AI-based predictions, offering a potential solution for identifying patients who may benefit from personalized therapeutic strategies, ultimately aiming "to improve the survival outcomes of GC patients."

Reference [7] Computational Pathology (CPath) is an interdisciplinary field focused on advancing computational approaches for the analysis and modeling of medical histopathology images. Developing digital diagnostics workflows and infrastructure is the main objective of CPath, which aims to revolutionize cancer diagnosis and treatment by acting as a clinical pathology computer-aided diagnosis (CAD) system. A paradigm shift in CPath has occurred with the recent explosion of deep learning as well as computer vision algorithms and the growing availability of digital pathology data. There is still a big divide in the application of these algorithms in clinical practice, even with the wealth of scientific and engineering advancements in the field of cancer image analysis. This review, which includes more than 700 papers, classifies each paper into a model card and tackles issues ranging from problem design to application. The thorough overview encompasses perspectives that are data-centric, model-centric, and application-centric, with the goal of assisting the community in comprehending the present state and potential future directions of CPath. The article's conclusion outlines the remaining difficulties and suggests future paths for computational pathology's continued advancement and clinical application. The readers are directed to GitHub for the most recent details regarding this survey review as well as access to the original model cards repository.

Reference [8] Lung cancer, with the highest global mortality rate among all cancers, underscores the critical importance of early detection and personalized treatment strategies to enhance the 5-year survival rate. Although manual means of evaluating medical images for lung cancer screening can be beneficial, there are drawbacks to using chest computed tomography (CT), such as the possibility of errors or misdiagnoses, as well as variations in physician interpretations. Lung cancer diagnosis and treatment have revolutionary new possibilities with the development of AI. AI makes extensive use of machine learning and deep learning technologies to perform tasks like subtype identification based on CT images, lung nodule detection, and

cancer case differentiation. AI also helps in non-invasively predicting genetic mutations and molecular status, which directs the best course of treatment and helps assess prognosis and therapeutic efficacy. By helping pathologists with molecular characterization, prognosis prediction, and typing, AI models based in histology improve the effectiveness of diagnosis and treatment. Despite the promising potential, widespread adoption in clinical settings encounters challenges such as data sharing, standardized labeling, regulatory considerations, and the integration of multimodal approaches. Nevertheless, AI stands as a promising catalyst in advancing lung cancer care.

## 5 Challenges Faced in Implementing Predictive Models in Real-World Pathology

### 5.1 Upstream Issues of AI

**Data Acquisition:** In order to train machine learning models, data must be gathered and prepared. One of the difficulties is obtaining high-quality, representative, and diverse data.

**Algorithm Training:** Choosing the right architectures, hyperparameters, and optimization strategies is a necessary step in creating efficient algorithms. It is essential to guarantee robustness, fairness, and transparency throughout training.

### 5.2 Downstream Issues of AI

**Dataset Shift:** Performance may suffer if training data is not identical to real-world data. Models must be updated for new situations.

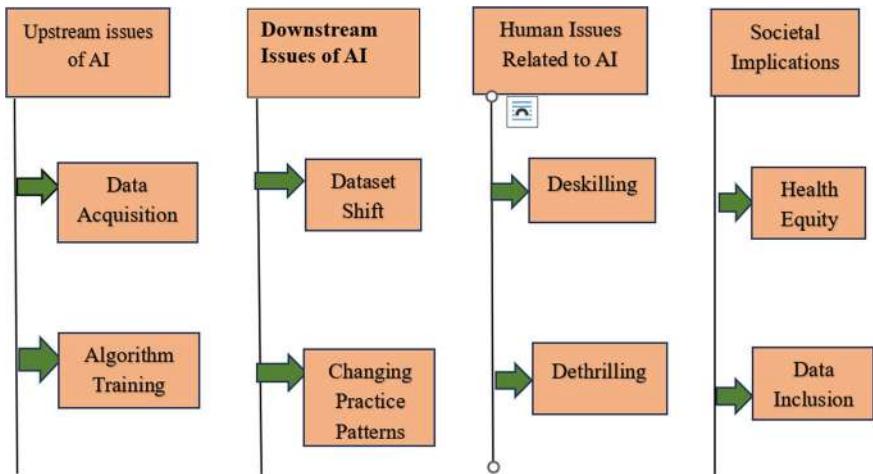
**Model Recalibration:** As surroundings change over time, models may drift. Accurate forecasts are ensured by routine recalibration.

**Modifying Practice Norms:** Adoption of AI has an impact on professional practices, decision-making, and processes. It is imperative to adjust to shifts induced by AI.

### 5.3 Human Issues Related to AI

**Deskilling:** Automation can make some skills less necessary, which could cause workers to become less skilled.

**Dethrilling:** Reliance on AI too much might take away from the excitement of innovation and problem-solving.



**Fig. 6** Challenges faced in implementing predictive models in real-world pathology settings

Burnout: Stressing out human resources due to high expectations and reliance on AI might result in burnout.

## 5.4 Societal Implications

Health Equity: The use of AI in applications shouldn't make gaps in health worse. It is essential to guarantee equal access and advantages.

Data Inclusion: Representativeness in training data impacts AI performance. Inclusive data collection is vital for equitable outcomes (Fig. 6).

## 6 Performance Metrics

### 6.1 Sensitivity and Specificity

#### Sensitivity (True Positive Rate)

A high sensitivity in the pathology context indicates that the model can correctly detect and categorize malignant areas in digital pathology images.

Calculated by dividing the total number of true positives (TP) by the total number of false negatives (FN), sensitivity is an important measure in the assessment process.

A high specificity in pathology indicates that the model can reliably discriminate between healthy and malignant tissue.

To calculate specificity, one uses the ratio of true negatives to the sum of true negatives and false positives (FP).

$$\text{Sensitivity} = \frac{\text{True Positives(TP)}}{\text{True Positives(TP)} + \text{False Negatives(FN)}}$$

## 6.2 Accuracy

The percentage of true positives and true negatives that are correctly predicted in relation to all of the instances in the dataset is known as precision.

It serves as an indicator of the predictive model's overall accuracy.

Accuracy provides an easy-to-understand metric of how well the model performs across all classes, but it may not be sufficient if there is a significant class imbalance.

The accuracy formula is given by:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Instances}}$$

## 6.3 Precision and Recall

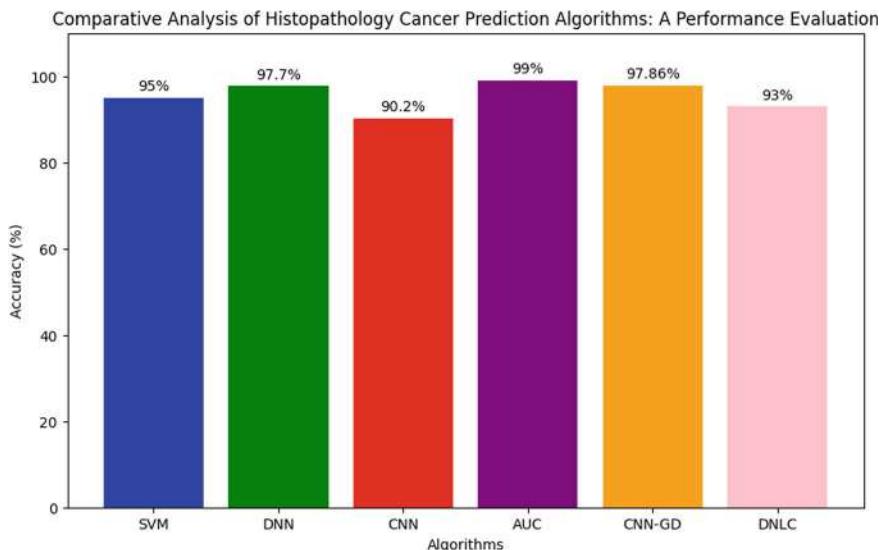
**Precision:** Precision, sometimes referred to as positive predictive value, measures how accurate the model is at predicting a positive outcome—in this case, cancer.

The ratio of true positives to the total of TP and false positives is employed to compute it.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

**Recall (Sensitivity):** Referred to as recall, sensitivity, or the TP rate, recall measures the model's ability to correctly identify positive cases (like cancer) out of all TP cases. True positives are divided by the sum of TP and FN in order to calculate it (Fig. 7).

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$



**Fig. 7** Comparative analysis of histopathology cancer prediction algorithms: A performance evaluation

## 7 Conclusion

In conclusion, our survey of the current landscape in pathology within the digital era reveals a compelling and rapidly evolving scenario, particularly in the realm of cancer detection. The comprehensive review of predictive models presented in this paper underscores the transformative potential of technology in reshaping traditional pathology practices. The field of digital pathology has made major advances in integrating AI and ML algorithms, which present previously unheard-of possibilities for enhancing the precision and effectiveness of cancer detection. The array of predictive models discussed in this survey highlights their diverse applications, from early diagnosis to personalized treatment strategies, fostering a paradigm shift toward more effective healthcare outcomes. However, as we navigate the path toward widespread implementation, it is imperative to acknowledge the challenges inherent in this transformative journey. Issues related to data privacy, the standardization of methodologies, and the imperative need for large and diverse datasets for robust model training pose substantial hurdles. In order to guarantee the accuracy and moral application of predictive models in pathology, it is imperative that these issues be resolved. While acknowledging the challenges, the immense potential for enhancing diagnostic accuracy and patient outcomes is evident.

## References

1. Santos T, Tariq A, Gichoya JW, Trivedi H, Banerjee I. Automatic classification of cancer pathology reports: a systematic review
2. Kather JN, Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT. A new generation of clinical biomarkers through deep learning in cancer pathology
3. Donovan MJ, Costa J, Cordon-Cardo C. Systems pathology a paradigm shift in the practice of diagnostic and predictive pathology
4. Song Z, Zou S, Zhou W, Huang Y, Shao L, Yuan J, Gou X, Jin W, Wang Z, Chen X, Ding X, Liu J, Yu C, Ku C, Liu C, Sun Z, Xu G, Wang Y, Zhang X, Wang D, Wang S, Xu W, Davis RC, Shi H. Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning
5. Vergheese G, Lennerz JK, Ruta D, Ng W, Thavaraj S, Siziopikou KP, Naidoo T, Rane S, Salgado R, Pinder SE, Grigoriadis A. Computational pathology in cancer diagnosis, prognosis, and prediction—present day and prospects
6. Huang B, Tianb S, Zhanc N, Mad J, Huange Z, Zhange C, Zhange H, Minge F, Liaoa F, Jia M, Zhanga J, Liua Y, Hea P, Denga B, Hua J, Donga W. Accurate diagnosis and prognosis prediction of gastric cancer using deep learning on digital pathological images: a retrospective multicentre study
7. Hosseini MS, Bejnordi BE, Trinh VQ-H, Chan L, Hasan D, Li X, Yang S, Kim T, Zhang H, Wu T, Chinniah K, Maghsoudlou S, Zhang R, Zhu J, Khaki S, Buin A, Chaji F, Salehi A, Nguyen BN, Samaras D, Plataniotis KN. Computational pathology: a survey review and the way forward
8. Shao J, Feng J, Li J, Liang S, Li W, Wang C. Novel tools for early diagnosis and precision treatment based on artificial intelligence

# Detection of Coal Miner with a Comprehensive Dataset Using Transfer Learning Techniques



**Md. Sazedur Rahman, Khandoker Hoque, Md. Boktiar Hossain, Denesh Das, and Tao Wu**

**Abstract** In underground mining sites, detecting coal miners is essential to protecting their safety and welfare. Nonetheless, there are many difficulties in this task. Significant obstacles include low visibility, complicated backgrounds, erratic environmental conditions, and a lack of annotated data. Human considerations and real-time processing requirements add to the task's complexity. To overcome these obstacles, multidisciplinary approaches utilizing cutting-edge sensor technologies, machine learning techniques, and domain-specific expertise are required. Various efforts have been made to detect the mine workers but suffer from several challenges like: Absence of real-time detection system, low detection accuracy. In this study, we have utilized on comprehensive dataset called DSLF+ to fine tune two cutting-edge detection model—DETR and YOLOv7 to accurately detect the coal miners. Intensive experiments have been conducted to show the effectiveness of the proposed method. YOLOv7 has shown a detection accuracy of 90% and DETR model has shown a maximum recall of 90%.

## 1 Introduction

Coal mining is experiencing a revolution as a result of the advancement of automated mining technology and sophisticated information technology [15]. In this context, the idea of “smart mining” has been put up to encourage environmentally friendly,

---

Md. S. Rahman (✉) · T. Wu

Department of Computer Science, Missouri University of Science and Technology, Rolla, MO, USA

e-mail: [mrvfw@umsystem.edu](mailto:mrvfw@umsystem.edu)

T. Wu

e-mail: [wuta@mst.edu](mailto:wuta@mst.edu)

K. Hoque · Md. B. Hossain

School of Engineering, San Francisco Bay University, Fremont, CA, USA

e-mail: [khoque43977@student.sfbu.edu](mailto:khoque43977@student.sfbu.edu)

D. Das

Department of Electrical and Computer Engineering, Lamar University, Beaumont, TX, USA

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2025

535

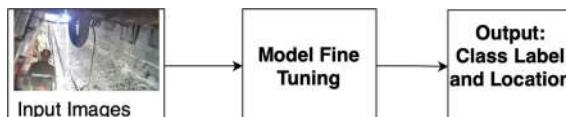
A. Kumar et al. (eds.), *Proceedings of Fourth International Conference on Computing and Communication Networks*, Lecture Notes in Networks and Systems 1292,

[https://doi.org/10.1007/978-981-96-3250-3\\_41](https://doi.org/10.1007/978-981-96-3250-3_41)

secure, and productive coal mining, and it has garnered a lot of interest from a variety of mining sectors [3]. Using specialist technology is essential for controlling coal mine safety in an efficient manner. Due to the continuous increase in monitoring data, abundance of incomplete data [14], traditional manual surveillance methods are no longer adequate to meet the ever-increasing safety standards. Researchers started using computer vision technology [6] to treat a range of visual problems as it advanced over time. Simultaneously, these technologies are being applied in the field of intelligent mines. The primary purpose of person search [23] is to ascertain if a particular individual is present in a picture or a video clip. Person search offers a lot of potential for use in video surveillance applications, such looking for suspects or lost persons. One of the real-life applications of person detection is in underground mines specifically in case of any accidents. In coal mines, person localization is used to quickly locate stranded miners and guarantee their safety in the event of an accident [25]. The complicated environment of coal mines makes underground people real-time position a hard problem even though it plays a significant role in coal mine output [7]. The safety protocols pertaining to miners entering and departing coal mines require that the quantity and types of miners entering and leaving the mines be determined. The two primary categories of convolutional neural network-based mainstream methods currently in use are one-stage and two-stage detection techniques. Two-stage detection techniques that are most frequently used are R-CNN [4], Faster R-CNN [19], Cascade R-CNN [1], and so on. The three most common one-stage detection techniques are the YOLO series [17, 18, 21, 22] and SSD [12], DETR [2]. The trade-off between detecting speed and precision is realized by this method. However, the recent works confront a number of difficulties when it comes to the target identification mission of underground coal mine staff. Like i) putting in place a simple system for people detection in real time and ii) increasing resilience to handle difficult environments. Figure 1 illustrates the system block diagram.

The major contributions of this works are:

- Transfer learning technique is correctly applied is computer vision tasks.
- One custom dataset is used to train the models to detect coal miners.
- Intensive experiments have been conducted to compare the performance of the fine-tuned models.



**Fig. 1** Proposed method

## 2 Related Works

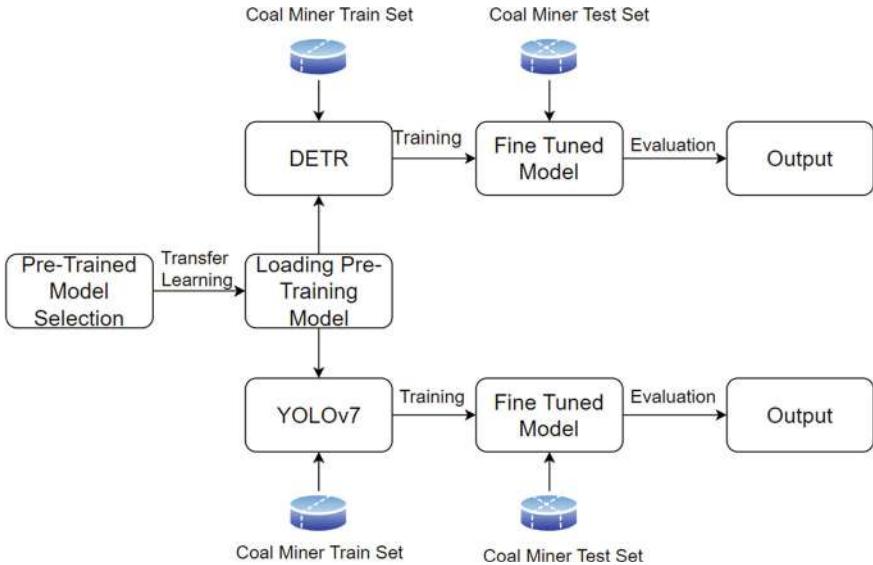
There have been several works related to object detection in computer vision by various researchers [16]. Hong et al. in [5] have suggested integrating the discrete wavelet transform (DWT) with the histogram of the oriented gradient (HOG) to create a pedestrian detection feature and method. The technique sets the region of interest (ROI) based on motion magnitude to increase detection speed. The ROI and Hough transform combination for persons detection was enhanced by Li et al. [13]. Numerous academics have tackled the subject of underground personnel detection using the most advanced detection algorithm using deep learning networks. In Li et al. in [9], to improve their resolutions, the infrared personnel photos are preprocessed using a super-resolution convolutional neural network technique. Secondly, the YOLOv4 network's detection performance is enhanced by optimizing the estimation of anchor boxes prior to network training by the application of K-means++ clustering. Based on the YOLOv5 technique, Kou et al. [8] suggested a lightweight Ucm-YOLOv5 algorithm that met the required real-time detection speed on the CPU side. In [26], the enhancing algorithm was applied to the low-light image. The Dense-YOLO method then identified the improved image, resolving the issue of mine staff detection being undetected. We suggest the YOLOv5-GS model, which has been trained using a dataset of people. The benefits of both the squeeze-and-excitation block and the ghostnet are combined in YOLOv5-GS, which speeds up human detection. Li et al. [10] have created a model adaption architecture for the feature extraction module that can choose multiple networks based on the population size. In order to accurately locate underground coal mine operators in challenging circumstances, shao et al. in [20] have introduced a new Re-parameterization YOLO (Rep-YOLO) detection technique.

## 3 Methodology

We have adopted the concept of transfer learning. The process of applying previously acquired knowledge from a related task to a new task to increase learning efficiency is known as transfer learning. Some of the most advanced object detection models are now on the market. We investigated the use of pre-trained models that have already been trained on the MS COCO dataset [11]. Figure 2 depicts the overall procedure.

### 3.1 Model Selection

A key component of our work is selecting the right pre-trained model to train using our special dataset. Considerations to make before selecting the finest pre-trained model are numerous. These are speed (ms), COCO mAP, and output type. The output types



**Fig. 2** Proposed method

represent yet another crucial matter. Bounding box output is the standard approach. On the basis of these, we have selected two state-of-the-art models.

- **DETR:** The DEtection TRansformer (DETR) [2] model uses a transformer-based architecture to process whole images in a single pass, thereby revolutionizing object detection. In contrast to conventional techniques, DETR predicts class labels and bounding boxes for every object at the same time, doing away with the requirement for intricate region proposal systems. DETR delivers remarkable performance on several datasets by approaching object detection as a set prediction issue and utilizing multi-head self-attention methods, so establishing a new benchmark for object detection tasks in terms of ease of use, effectiveness, and efficiency.
- **YOLOv7:** “You Only Look Once” version 7, or “YOLOv7,” [21] is a noteworthy development in real-time object identification. YOLOv7 improves accuracy and speed over its predecessors by enhancing computational efficiency and fine-tuning the network design. YOLOv7 processes pictures quickly with a single step technique, anticipating bounding boxes and class probabilities without requiring complex region proposal procedures. Because of its simplified procedure, YOLOv7 can detect objects in a variety of situations quickly and accurately. This makes it an effective tool for applications like robotics, autonomous driving, and surveillance that need real-time object recognition.

Table 1 displays the performance results of DETR on COCO 2017 dataset.

**Table 1** DETR model performance and details

Name	Backbone	Box AP	Segm AP	PQ	Size (Mb)
DETR	R50	38.8	31.1	43.4	165
DETR-DC5	R50	40.2	31.9	44.6	165
DETR	R101	40.1	33	45.1	237

**Table 2** Model parameters for YOLOv7

Name	Value
Layers	425
Parameters	37,196,556
Gradients	37,196,556
GFLOPS	105.1
Weight decay	0.0005
Batch size	4
Epochs	10
Optimizer	Adam
Momentum	0.937
Learning rate	0.01

### 3.2 Model Training

Selecting the suitable models is the first step toward the challenge of training the models. The two primary processes in training a model are splitting the dataset and selecting the model parameters.

1. *Dataset Split:* We have total 30706 of coal miner class. Among which 24564 images are used for training and 6141 images are used for validation. The train, validation, and test sets of the dataset are split up in an 80:20 ratio.
2. *Model Parameter Selection:* An essential component of model training is parameters. If a model's parameters are given inconsistently or incorrectly, it may perform poorly. Two separate settings have been used for DETR and YOLOv7 because of their inherent differences in architecture. Table 2 displays the model summary and parameters used for training. On the other hand, Table 3 shows the parameters used for training DETR model.

## 4 Experiments

It takes a significant amount of processing power and computational resources to train a pre-trained model. The experimental setup is shown below.

**Table 3** Model parameters for DETR

Name	Value
Backbone	resnet50
Batch size	8
Drop out	0.1
Epochs	10
Learning rate	0.001
Encoder layers	6
Hidden layer dimension	256
Number of heads	8
Number of workers	2
Weight decay	0.0001
Parameters	41,302,368

## 4.1 Experimental Setup

Experiments are conducted in Almalinux v8.9 setup with 8 CPU cores, NVIDIA A100 80 GB GPU. For training DETR model, we have used CUDA version 11.8, Python 3.8 and PyTorch version 2.3.0. And for training YOLOv7 model, we have used version 11.1, Python 3.8.19 and PyTorch version 1.9.0.

## 4.2 Dataset

We utilize an image dataset of underground longwall mining face called DsLMF+ which is proposed in [24]. There are 138,004 photos total, divided into six groups: huge coal, towline, miners' conduct, mine personnel, hydraulic support guard plate, and mine safety helmet. In order to make the dataset labels compatible with widely used target detection networks, they are provided in both YOLO and COCO formats. Photographs from completely automated coal mining faces in Shaanxi Province, China, were gathered from original underground monitoring footage. IVG-G5A network HD cameras and Openmv IMX335 lenses were among the specialized equipment used to process the photographs. As our main focus was to detect the mine personnel, so we have used the “coal-miner” class. This class contains total 30,706 images among which 24,564 images are used for training and 6141 images are used for validation. Figure 3 illustrates some of the images of coal miners form the dataset.



**Fig. 3** Sample photos from the DsLMF+ dataset featuring the coal miner class [24]

### 4.3 Evaluation of YOLOv7

In this section, we provide different evaluation metrics to justify the performance of the fine-tuned YOLOv7 model. Figure 4 shows some example of prediction made by YOLOv7 model. It is seen that coal miners are detected accurately with correct bounding box around them.

The recall and precision of the YOLOv7 model are shown in Fig. 5. It is evident that both of these values grow as the number of epochs increases.

Figure 6a displays the mean average precision at a threshold of 0.5 for intersections over unions (IoU), and in Fig. 6b, mean average precision at an intersection over union (IoU) from of 0.5 to 0.9 is shown.

A term used to describe the loss function used in training to determine whether an item is present in a given bounding box is “objectness loss.” In object detection models, “box loss” usually refers to the loss function connected to the bounding box regression component. Figure 7a, b illustrates the bounding box loss and objectness loss of the model as the epoch increases.

Lastly in Fig. 8, the F1-curve is shown of fined tuned YOLOv7 model.

### 4.4 Evaluation of DETR

In this section, the evaluation of DETR model is described. Figure 9 shows the cross-entropy loss that decreases with the epochs.

In Fig. 10, the bounding box loss is illustrated. When performing object identification tasks, the generalized intersection over union (GIoU) loss function is frequently employed in conjunction with anchor-based object recognition models like you only look once (YOLO) or Faster R-CNN. In Fig. 11 the GIoU loss is displayed.

**key observations:** While training the DETR and YOLOv7 model, it is observed that YOLOv7 performed better than DETR model. The problem with DETR model is that it has encoder-decoder architecture. Moreover the original model is trained on distributed GPUs with larger batch size but we have used only one GPU with lesser batch size. Improving the performance of DETR is taken as a future extension of our work.

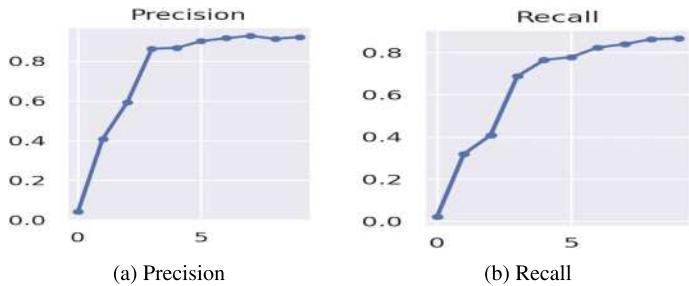
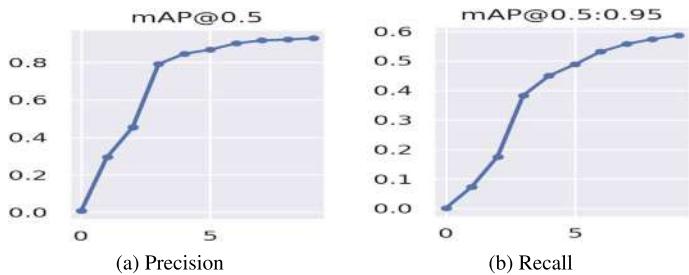
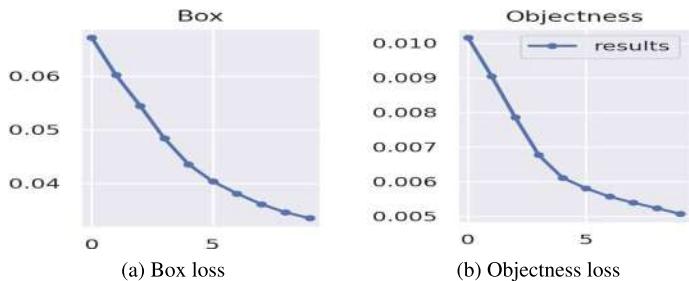


**Fig. 4** Bounding box prediction of YOLOv7 model

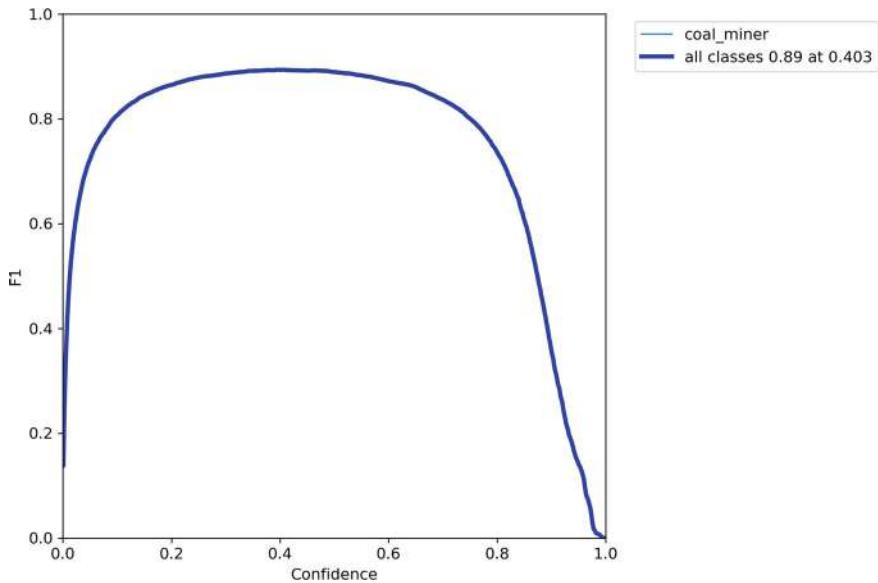
Lastly, Tables 4 and 5 illustrate the evaluation performance on the new dataset. Evaluation is conducted on several values of IoU and area.

## 5 Conclusion and Future Work

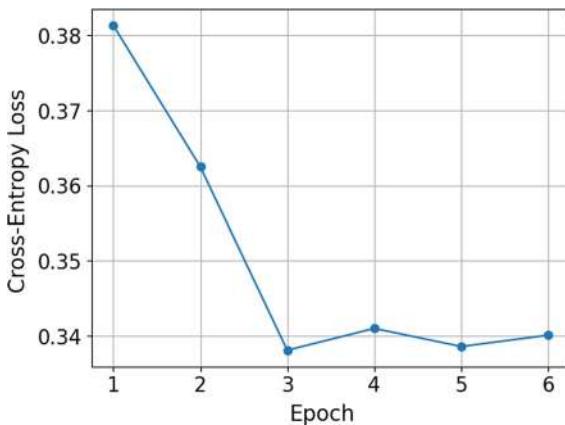
With the advancement of automation, more better technological solution is required in all fields. In this paper, we have proposed two finned tuned model—DETR and YOLOv7 to correctly detect the coal miners. We have trained the models using the DsLMF+ dataset that contains a huge collection of images collected from under-

**Fig. 5** Precision and recall of fine-tuned YOLOv7 model**Fig. 6** Mean average precision of fine-tuned YOLOv7 model**Fig. 7** Box loss and objectness loss of fine-tuned YOLOv7 model

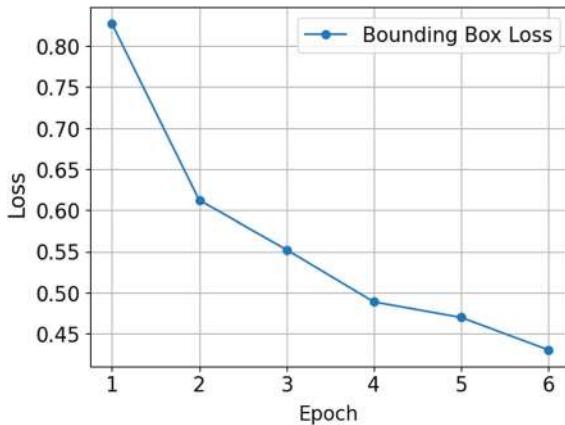
ground mines. This work can show the path for better monitoring and disaster management systems in underground mines. Despite a good performance, there is still room for improvement. Specially, for DETR model, we need to normalize the images for better accuracy as the original DETR model is trained on Imagenet dataset. Also due to resource constraint, we could not train on huge number of epochs. In the future, we can train the model on larger parameter settings to generate better performance. Moreover, we plan to employ ensemble technique to accumulate the individual prediction of the model.



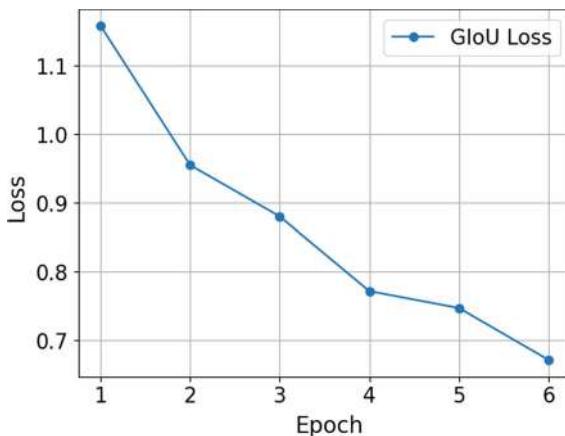
**Fig. 8** F1-curve of fine-tuned YOLOv7 model



**Fig. 9** Epochs versus cross-entropy loss



**Fig. 10** Epochs versus bounding box loss



**Fig. 11** Epochs versus GIoU loss

**Table 4** Average precision on DETR model

Mean average precision		
IoU = 0.50:0.95 Area = all	IoU = 0.50 Area = all	IoU = 0.50:0.95 Area = large
0.003	0.011	0.003

**Table 5** Recall on DETR model

Metrics	IoU = 0.50:0.95 Area = All MaxDets = 1	IoU = 0.50:0.95 Area = All MaxDets = 10	IoU = 0.50:0.95 Area = All MaxDets = 100	IoU = 0.50:0.95 Area = Medium MaxDets = 100	IoU = 0.5:0.95 Area = Small MaxDets = 100
Value	0.068	0.073	0.075	0.036	0.090

## References

1. Cai Z, Vasconcelos N (2018) Cascade R-CNN: delving into high quality object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 6154–6162
2. Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S (2020) End-to-end object detection with transformers. In: European conference on computer vision. Springer, , pp 213–229
3. Ge X, Shuai S, Haiyang Y, Chen G, Xiaoping L (2018) Smart mine construction based on knowledge engineering and internet of things. Int J Perform Eng 14(5):1060
4. Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 580–587
5. Hong G-S, Kim B-G, Hwang Y-S, Kwon K-K (2016) Fast multi-feature pedestrian detection algorithm based on histogram of oriented gradient using discrete wavelet transform. Multimed Tools Appl 75:15229–15245
6. Huang TS (1996) Computer vision: evolution and promise. In: CERN European organization for nuclear research-reports-CERN, pp 21–26
7. Jiaxi W, Yan C, Shuo S (2011) An improved TDOA algorithm applied person localization system in coal mine. In: 2011 Third international conference on measuring technology and mechatronics automation, vol 1. IEEE, pp 428–431
8. Kou F, Xiao W, He H, Chen R (2023) Research on target detection in underground coal mines based on improved YOLOv5. 45(7):2642–2649
9. Li X, Wang S, Liu B, Chen W, Fan W, Tian Z (2022) Improved YOLOv4 network using infrared images for personnel detection in coal mines. J Electron Imaging 31(1):013017–013017
10. Li Y, Yin K, Liang J, Tan Z, Wang X, Yin G, Wang Z (2023) A multitask joint framework for real-time person search. Multimed Syst 29(1):211–222
11. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Lawrence Zitnick C (2014) Microsoft coco: Common objects in context. In: Computer vision–ECCV 2014: 13th European conference, Zurich, Switzerland, 6–12 Sept 2014, Proceedings, Part V 13. Springer, , pp 740–755
12. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC (2016) SSD: Single shot multibox detector. In: Computer vision–ECCV 2016: 14th European conference, Amsterdam, The Netherlands, 11–14 Oct 2016, Proceedings, Part I 14. Springer, pp 21–37
13. Ning Z, Shanjun M, Mei L (2017) Enhancement algorithm based on illumination adjustment for non-uniform illuminance video images in coal mine. J China Coal Soc (8)
14. Rahman Md S, Azharul Hasan KM (2023) Computing skyline query on incomplete data. In: International conference on big data, IoT and machine learning Springer, pp 657–672
15. Rahman Md S, Elmahallawy M, Madria S, Frimpong S (2024) CAV-AD: a robust framework for detection of anomalous data and malicious sensors in CAV networks. [arXiv:2407.05461](https://arxiv.org/abs/2407.05461)
16. Rahman Md S, Hassan Md Z, Hossain SN, Masrur N, Rabbi J (2022) Bangladeshi local vehicle recognition with a comprehensive dataset using transfer learning techniques. In: 2022 4th International conference on sustainable technologies for industry 4.0 (STI). IEEE, pp 1–6
17. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 779–788

18. Redmon J, Farhadi A (2017) Yolo9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7263–7271
19. Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in neural information processing systems, vol 28
20. Shao X, Liu S, Li X, Lyu Z, Li H (2024) Rep-YOLO: an efficient detection method for mine personnel. *J Real-Time Image Process* 21(2):1–16
21. Wang C-Y, Bochkovskiy A, Mark Liao H-Y (2023) 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, pp 7464–7475
22. Xu S, Wang X, Lv W, Chang Q, Cui C, Deng K, Wang G, Dang Q, Wei S, Du Y et al (2022) PP-YOLOE: an evolved version of YOLO. [arXiv:2203.16250](https://arxiv.org/abs/2203.16250)
23. Xu Y, Ma B, Huang R, Lin L (2014) Person search in a scene by jointly modeling people commonness and person uniqueness. In: Proceedings of the 22nd ACM international conference on multimedia, pp 937–940
24. Yang W, Zhang X, Ma B, Yujia W, Yan J, Liu Y, Zhang C, Wan J, Wang Y, Wang Y et al (2023) An open dataset for intelligent recognition and classification of abnormal condition in longwall mining. *Sci Data* 10(1):416
25. Yang Y, Li Y, Guo X (2012) Underground personnel positioning system based on low-power card reader. In: International conference on automatic control and artificial intelligence (ACAI 2012). IET, pp 2239–2242
26. Zhang MZ (2022) Underground pedestrian detection model based on dense-yolo network. *Ind Mining Autom* 48:86–90

# A Method for Generating a Computer Simulation Model for Detecting Scoliosis Images



Yijun Zhang, Huanxiang Ding, and Jifeng Zhou

**Abstract** Scoliosis can affect the patient's physical appearance, cause pain, and cause difficulty breathing. With the continuous progress of medical technology and the continuous development of image processing technology, the accuracy of early detection and diagnosis of this disease has been improved. Therefore, this article aims to apply computer simulation technology to detect scoliosis images using models and improve the level of early intervention. This article mainly designs a detection model and generates images of scoliosis using computer simulation technology based on experimental and comparative methods, tests the performance of four models, compares the accuracy, recall, and F1 value of the selected Faster Region Convolutional Neural Networks (Faster R-CNN), analyzes the image generation results of the training, validation, and testing sets, and explores the capabilities of the three image generation techniques. The experimental results show that when the learning rate is 0.001, the Faster R-CNN model has the highest classification accuracy (0.92). Image transformation technology has strong robustness in generating scoliosis structures.

**Keywords** Computer simulation · Scoliosis · Image detection · Model generation

## 1 Introduction

The incidence of scoliosis is increasing year by year globally, with a more prominent issue among the adolescent population. Early detection and treatment of scoliosis are of great significance in reducing patient pain and preventing the deterioration of the condition. The early symptoms of scoliosis are not obvious, leading to many patients

---

Y. Zhang · H. Ding · J. Zhou (✉)

School of Physical Education and Health, Linyi University, Linyi 276000, Shandong Province, China

e-mail: [zhjf8882003@126.com](mailto:zhjf8882003@126.com)

H. Ding

e-mail: [dinghuanxiang@lyu.edu.cn](mailto:dinghuanxiang@lyu.edu.cn)

being diagnosed only after their condition becomes severe. Deep learning algorithms can automatically extract valuable features for diagnosis by learning feature information from a large amount of data, improving the accuracy and efficiency of diagnosis.

Computer simulation technology can accurately calculate the curvature, angle, and displacement of the spine, and quickly process and interpret image data of scoliosis. Its high degree of automation and standardization reduces the impact of human error. The three-dimensional reconstruction of image data allows doctors to have a more intuitive understanding of the spinal condition. Interactive operations help to observe changes in the spine. This technology also has its place in auxiliary teaching and training for scoliosis.

This article first discusses the incidence of scoliosis and analyzes the advantages of computer simulation. Secondly, this article provides a review of the theories of scoliosis studied by other scholars. In the third part, this article studies the image detection model for scoliosis, mentions image processing and generation techniques, briefly introduces scoliosis, and analyzes several classification detection models. In the fourth part, this article conducts experimental design and result analysis on the detection model, and obtains relevant data results through three sets of experiments. Finally, this article provides a summary and suggestions.

## 2 Related Works

This article aims to construct an efficient and accurate image detection model for scoliosis by conducting in-depth research on image processing techniques, machine learning algorithms, and deep learning algorithms. Therefore, this article first discusses the research on scoliosis by some scholars. Ruiz G conducted a comprehensive narrative review of early onset scoliosis. He discussed the definition, classification, etiology, diagnosis, treatment methods, and prognosis of early onset scoliosis, emphasizing the importance of early diagnosis and intervention, and proposing management strategies for this type of disease [1]. Lee G B paid special attention to the different types, causes, and current treatment methods of adolescent scoliosis, providing readers with comprehensive knowledge of scoliosis [2]. Based on spinal segmental assessment, Khodjayeva DI conducted a morphological study on idiopathic scoliosis, which detailed the different types and degrees of scoliosis [3]. Marya S analyzed the multifactorial etiology theory of adolescent idiopathic scoliosis and explored the roles of genetic, environmental, and biomechanical factors in disease occurrence [4]. Through system review and meta-analysis, Dimitrijević V evaluated the impact of Schroth's method and core stabilization exercises on idiopathic scoliosis, providing information on the effectiveness of non-surgical treatment for idiopathic scoliosis [5]. Glavaš J believes that school medicine can play a role in the early detection and management of adolescent idiopathic scoliosis [6]. Negrini S aimed to provide a unified classification system for corrective devices for scoliosis. This classification system covers different types of correctors and their indications

[7]. Ashebo L had updated the latest knowledge on the diagnosis and management of early onset scoliosis [8]. To compare the efficacy and safety of these two treatment methods for adolescent idiopathic scoliosis patients, Charalampidis conducted a randomized clinical trial to explore the effects of nighttime orthosis treatment and exercise therapy [9]. Sebaaly systematically analyzed and summarized the treatment methods for congenital scoliosis, and proposed an evidence-based treatment algorithm [10]. To evaluate the lung function of untreated children and adolescents with idiopathic scoliosis, Kempen applied meta regression analysis to explore the relevant influencing factors [11]. Gargano et al. investigated the relationship between melatonin and adolescent idiopathic scoliosis. They reviewed existing research, evaluated the therapeutic effects and mechanisms of melatonin in adolescent idiopathic scoliosis patients, and made recommendations for future research directions [12]. Ormonjonovich explored the accuracy and reliability of different methods in determining the degree of scoliosis [13]. By measuring and analyzing the trunk morphological parameters of children with scoliosis, Muzafarovna indicated that scoliosis had a significant impact on the patient's body structure [14]. Motye conducted qualitative research and conducted interviews with patients and their families to understand their psychological feelings and concerns before surgery, and explored the psychological experience of adolescent idiopathic scoliosis patients before surgery [15]. Therefore, this article delves into computer simulation technology and detection model generation in depth.

### 3 Spinal Scoliosis Image Detection Model

#### 3.1 Image Processing Technology

Image processing technology involves image digitization, feature extraction, image enhancement, segmentation, recognition, compression, synthesis, analysis, repair and reconstruction, and three-dimensional operations. It is widely used in fields such as medical image analysis, autonomous driving, security monitoring, artistic creation, social media, and entertainment industry.

Median filtering can effectively remove salt and pepper noise in images and improve image clarity. Histogram equalization enhances the contrast of the image. In this technology, feature extraction is performed using edge detection, texture analysis, and morphological processing. Canny edge detection accurately extracts edge information of scoliosis. This article uses Gaussian filtering to achieve smoothing, and its formula is:

$$f(a, b) = 1/(2\pi\mu^2) * e^{-((a^2+b^2)/(2\mu^2))} \quad (1)$$

(a, b) are the coordinates of the relative center, and  $\mu$  is the standard deviation. Then using the Sobel operator to calculate the first derivative in the horizontal and

vertical directions, and calculate the gradient amplitude and direction. The amplitude of the gradient is expressed as:

$$G = \sqrt{Ga^2 + Gb^2} \quad (2)$$

The direction of gradient  $\theta$  is:

$$\theta = \arctan(Gb/Ga) \quad (3)$$

Scanning the entire image to remove non boundary points, with the aim of determining whether each pixel is a non-boundary point and suppressing it. Grayscale co-occurrence matrix is used to obtain texture features of scoliosis. Combining image processing technology with deep learning algorithms to achieve automatic recognition and classification of scoliosis images.

Convolutional neural networks simulate the connectivity of human brain neurons and can automatically extract features from images, classify and recognize them. It can extract the morphology and structural information of the spine from the image to determine whether there is a phenomenon of scoliosis. Support Vector Machine searches for the optimal hyperplane partition data to maximize the spacing between different categories of data. Random forest constructs multiple decision trees and integrates their output results. The application of these algorithms in image detection of scoliosis not only improves the accuracy of detection, but also reduces the dependence on artificial experience. Recurrent neural networks capture temporal information in scoliosis images to more accurately determine the degree of scoliosis, utilize generative adversarial networks to obtain more scoliosis images, and expand the dataset.

### 3.2 Scoliosis

One common symptom of scoliosis is a significant increase in one shoulder [16, 17]. If there is lateral curvature in the thoracic section of the spine, it can cause asymmetry in the shoulder lines on both sides [18, 19]. If it occurs in the lumbar region, there will be a situation where one end of the waistline increases and the other end disappears [20, 21]. The external manifestation of asymmetry in the hip joint or pelvis is unequal leg length. Lower limb weakness and paralysis are usually symptoms only present in severely ill patients. Congenital, idiopathic, or scoliosis caused by other diseases such as neurofibromatosis should be treated conservatively (wearing braces, regular observation, and doing back muscle exercises to delay the progression of scoliosis) [22, 23]. The degree of lateral curvature can greatly affect the development of cardiovascular function in patients and requires correction through surgery. Scoliosis can bring inconvenience to people's daily lives. Therefore, this article intends to detect and generate scoliosis structures through computer simulation, and intervene and deal with such diseases in a timely manner.

### 3.3 *Image Generation Technology*

Three-dimensional modeling can accurately simulate the shape and features of scoliosis, generating highly realistic images. A high-precision three-dimensional scanner obtains geometric data of the spine, processes this data, and constructs a three-dimensional model of the spine. It can also simulate different degrees of scoliosis by adjusting model parameters. Image transformation technology allows for rotation, scaling, translation, and other operations to simulate the morphological changes of scoliosis while maintaining the basic structure of the image. The advantage of this method is that it can quickly generate simulated scoliosis images without increasing actual shooting costs. This article first collected a batch of X-ray images of normal spine as reference data. Then, through image transformation, setting the angle of scoliosis from 5 to 40 degrees, and generate a set of images every 5 degrees. The common data augmentation methods are rotating and flipping images. This approach increases the rotation invariance and mirror invariance of the model. Brightness adjustment, contrast adjustment, and noise addition can also simulate changes in actual shooting conditions, making the model more robust. Autoencoder is an unsupervised learning model that learns low dimensional representations of input data. This article encodes the original scoliosis image into a low-dimensional feature representation, and then decodes it into an enhanced image.

### 3.4 *Analysis of Detection Models*

This article will annotate samples of scoliosis from different angles and severity, and label normal and abnormal spinal positions. This article selects Inception, EfficientNet, Faster Region Convolutional Neural Networks (Faster R-CNN), and You Only Look Once (YOLO) classifiers as detection models. The performance of these classification models at different learning rates is shown in Table 1.

This article sets the learning rate to 0.001–0.009, and the model can converge faster. The classification accuracy of the Inception model remains between 80–90%, and there is not much difference between EfficientNet and Inception. However, compared to Inception, the Inception model performs better. The Faster R-CNN and YOLO models perform better than the previous two, with the Faster R-CNN

**Table 1** Performance of classification models

Model	0.001	0.005	0.009
Inception	0.86	0.85	0.83
EfficientNet	0.88	0.84	0.81
Faster R-CNN	0.92	0.88	0.85
YOLO	0.90	0.87	0.84

model performing the best. Therefore, this article chooses the Faster R-CNN model for classification detection.

This article conducts supervised learning on the model, updates network weights, minimizes loss functions, and deploys the trained model to scoliosis images for real-time detection. This article uses regularization techniques to prevent overfitting of the model. This article introduces residual connections to improve the deep learning ability of the model in terms of architecture optimization. Residual connections allow the model to skip some layers during the training process, alleviating the problem of vanishing gradients.

## 4 Experimental Design and Result Analysis

### 4.1 Selection and Processing of Experimental Datasets

In order to improve the early detection rate and diagnostic accuracy of scoliosis, this article believes that advanced three-dimensional modeling, image transformation, and data augmentation technologies can generate simulated images with scoliosis characteristics. Model construction requires the application of deep learning algorithms, adjusting model parameters, and optimizing network structure. This article trains and tests the model using a dataset, compares it with other advanced algorithms, and evaluates the performance of the model. This article selected image datasets with various types of scoliosis and varying degrees of lesions to ensure the model's generalization ability and robustness. Congenital scoliosis usually occurs in childhood and is related to structural abnormalities in the spine. Neuromuscular scoliosis is caused by spinal muscle weakness and neurological problems. When the joints of the spine degrade, there is also a possibility of scoliosis occurring. Idiopathic scoliosis is the most common type, which occurs in adolescence. The gender difference is obvious. The incidence rate of women is much higher than that of men. The human spine consists of five parts: Cervical spine, thoracic spine, lumbar spine, sacral spine, and coccyx. There are seven cervical vertebrae, twelve thoracic vertebrae, and five lumbar vertebrae. These vertebrae are tightly connected through intervertebral disks, ligaments, and joints, forming a stable structure that supports the trunk and protects the spinal cord and nerve tissue. The spine deviates from the center position and bends to one side, becoming a skeletal deformity. The approximate structure of scoliosis is shown in Fig. 1.

These images cover patients with scoliosis of different genders and degrees of disease. This article uses preprocessing techniques to improve image quality and enhance the training effectiveness of the model. This article standardizes and normalizes images to eliminate the impact of factors such as lighting and contrast on image quality. Then, this article utilizes data augmentation techniques to increase the number of training samples and improve the model's generalization ability through



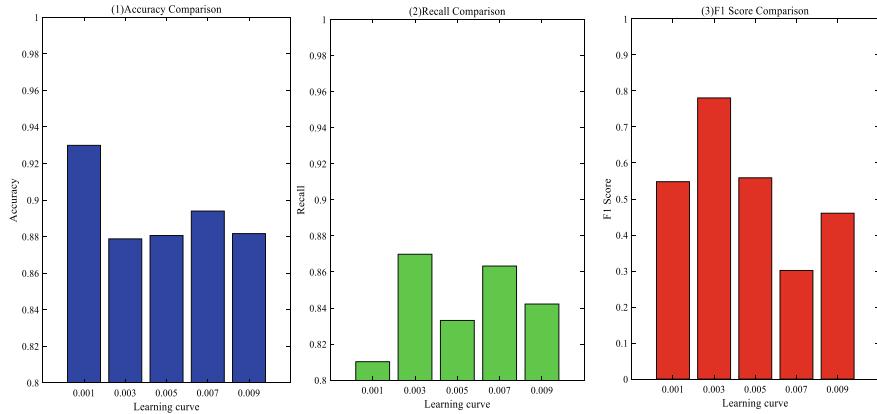
**Fig. 1** Spinal scoliosis structure

rotation, scaling, and translation. After increasing the number of samples, semi supervised learning is used to pre-train partially unlabeled image data. Extracting features from unsupervised learning algorithms and cluster unlabeled images to obtain initial category labels, which are then used as pseudo labels. Stratified sampling divides the dataset to ensure that the sample proportions of each category in the training, validation, and testing sets are similar, avoiding overfitting of the model.

## 4.2 Environmental Setup and Configuration

In terms of hardware, this article uses high-performance computer hardware, with Intel i7 multi-core processor and 128 GB of memory capacity. The high-speed Solid State Disk (SSD) uses SATA Express SSD and PCIe SSD. Graphics processors are referred to as display cores, visual processors, and display chips, and NVIDIA was chosen in this article. On the software side, this article will use the latest versions of deep learning frameworks TensorFlow and PyTorch as technical support, providing rich algorithm libraries and efficient computational performance. The image processing library uses Open Source Computer Vision Library and Python Imaging Library.

This article comprehensively evaluates the performance of the model using accuracy, recall, and F1 score as evaluation indicators. Firstly, the accuracy, recall, and F1 value of the image detection model for scoliosis were compared. Secondly, this article compared the effectiveness of the generated results for the training set, validation set, and test set images based on their loss values and errors. Finally, this article tested and analyzed the robustness and timeliness of three-dimensional modeling, image transformation, and data augmentation techniques in generating scoliosis images.



**Fig. 2** Comparison of accuracy, recall, and F1 values of image detection models for scoliosis

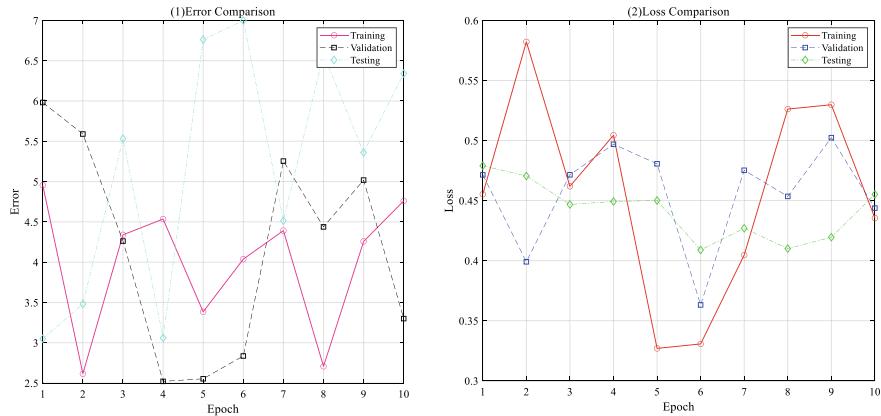
### 4.3 Experimental Results and Analysis

As shown in Fig. 2, this article finds that there are differences in the accuracy, recall, and F1 value performance of the detection model under the learning rates of 0.001, 0.003, 0.005, 0.007, and 0.009, respectively. Among them, the accuracy is highest and the recall is lowest when the learning rate is 0.001. When the learning rate is 0.007, the accuracy and recall rank second, with the lowest F1 value. The accuracy is the lowest, the recall is the highest, and the F1 value is the highest when the learning rate is 0.003.

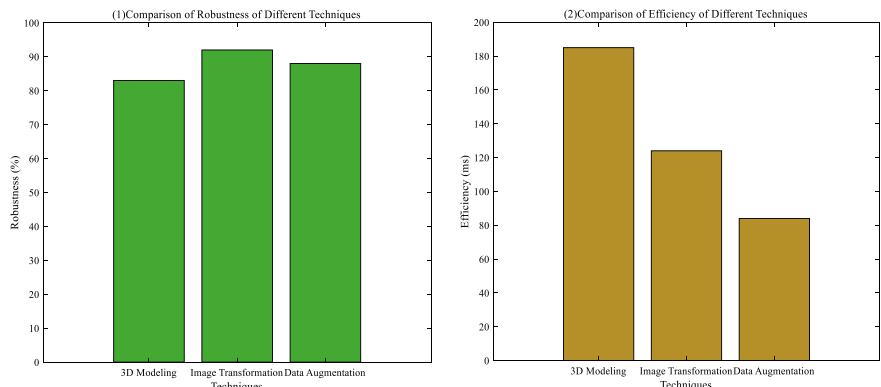
As shown in Fig. 3, this article can observe that the error and loss results of the training set, validation set, and test set are inconsistent at each stage. Among them, the maximum error value of the training set is 5%, the minimum is less than 3%, the maximum loss value exceeds 0.55, and the minimum is less than 0.35. The maximum error value of the validation set is 6%, the minimum value is close to 2.5%, the maximum loss value is 0.5, and the minimum is less than 0.4. The maximum error of the test set is 7%, the minimum is close to 3%, and the maximum loss value is between 0.45 and 0.5, and the minimum is close to 0.4.

This article compares the robustness and efficiency of three different technologies, and the specific results are shown in Fig. 4. Image conversion simulates the impact of noise on the original three-dimensional model, adding Gaussian noise with an average value of 0 and a standard deviation of 0.5 to the original data.

Figure 4 (1) shows the robustness comparison of three techniques, while Fig. 4 (2) shows the timeliness comparison. This article finds that image transformation technology has the strongest robustness (92%) and average timeliness (124 ms). The timeliness (185 ms) and robustness of three-dimensional modeling are the worst (83%). The robustness of data augmentation technology is moderate (88%), with the best timeliness (84 ms).



**Fig. 3** Image generation results for training, validation, and testing sets



**Fig. 4** Comparison of robustness and timeliness of three technologies

## 5 Conclusion

Scoliosis, also known as scoliosis, is a three-dimensional spinal deformity in which several segments of the spine bend laterally, accompanied by vertebral rotation. This situation is caused by various factors such as genetics, improper posture, inflammation, and tumors. The images of scoliosis generated through computer simulation can provide us with rich and diverse data resources, providing strong support for model training and optimization. This article applies experimental testing to compare the capabilities of three-dimensional modeling, image transformation, and data augmentation in image generation. Error analysis is also conducted on the results generated by the test set, validation set, and training set. The results indicate that the Faster R-CNN model can be used to detect scoliosis. Due to the large loss value, these three

datasets need to adjust and update the distorted data to expand the data volume in order to obtain more accurate numerical results.

**Acknowledgements** Fund Project: 2023 National Education Science planning key project of the Ministry of Education, Precise Intervention Research on spinal health promotion of children and adolescents from the perspective of Physical and Medical Integration, DLA230380.

## References

1. Ruiz G, Torres-Lugo NJ, Marrero-Ortiz P et al (2022) Early-onset scoliosis: a narrative review[J]. EFORT Open Rev 7(8):599–610
2. Lee GB, Priefer DT, Priefer R (2022) Scoliosis: causes and treatments[J]. Adolescents 2(2):220–234
3. Khodjayeva DI (2022) Morphology of idiopathic scoliosis based on segment by segment assessment of spinal column deformity[J]. Sci Prog 3(1):208–215
4. Marya S, Tambe AD, Millner PA et al (2022) Adolescent idiopathic scoliosis: a review of aetiological theories of a multifactorial disease[J]. Bone Joint J 104(8):915–921
5. Dimitrijević V, Viduka D, Šćepanović T et al (2022) Effects of Schroth method and core stabilization exercises on idiopathic scoliosis: a systematic review and meta-analysis[J]. Eur Spine J 31(12):3500–3511
6. Glavaš J, Rumboldt M, Karin Ž et al (2023) The role of school medicine in the early detection and management of adolescent idiopathic scoliosis[J]. Wien Klin Wochenschr 135(11):273–281
7. Negrini S, Aulisa AG, Cerny P et al (2022) The classification of scoliosis braces developed by SOSORT with SRS, ISPO, and POSNA and approved by ESPRM[J]. Eur Spine J 31(4):980–989
8. Ashebo L, Anari JB, Cahill PJ (2023) Update on the diagnosis and management of early-onset scoliosis[J]. Curr Rev Musculoskelet Med 16(10):447–456
9. Charalampidis A, Diarbakerli E, Dufvenberg M et al (2024) Nighttime bracing or exercise in moderate-grade adolescent idiopathic scoliosis: a randomized clinical trial[J]. JAMA Netw Open 7(1):e2352492–e2352492
10. Sebaaly A, Daher M, Salameh B et al (2022) Congenital scoliosis: a narrative review and proposal of a treatment algorithm[J]. EFORT Open Rev 7(5):318–327
11. Kempen DHR, Heemskerk JL, Kaçmaz G et al (2022) Pulmonary function in children and adolescents with untreated idiopathic scoliosis: a systematic review with meta-regression analysis[J]. Spine J 22(7):1178–1190
12. Gargano G, Oliva F, Migliorini F et al (2022) Melatonin and adolescent idiopathic scoliosis: the present evidence[J]. The Surgeon 20(6):e315–e321
13. Ormonjonovich IF (2022) Methods of determining degrees of scoliosis[J]. Asia Pacific J Market Manage Rev Impact Factor: 7.603 11(12): 319–324. ISSN: 2319-2836
14. Muzaferovna KS, Radjabovich BR, Joraboy S (2022) Morphometric parameters of the trunk in children with scoliosis[J]. Central Asian J Med Nat Sci 3(3):144–147
15. Motyer GS, Kiely PJ, Fitzgerald A (2022) Adolescents experiences of idiopathic scoliosis in the pre-surgical period: a qualitative study[J]. J Pediatr Psychol 47(2):225–235
16. Rebello D, Wohler E, Erfani V et al (2023) COL11A2 as a candidate gene for vertebral malformations and congenital scoliosis[J]. Hum Mol Genet 32(19):2913–2928
17. Gámiz-Bermúdez F, Obrero-Gaitán E, Zagalaz-Anula N et al (2022) Corrective exercise-based therapy for adolescent idiopathic scoliosis: systematic review and meta-analysis[J]. Clin Rehabil 36(5):597–608
18. Mens RH, Bisseling P, de Kleuver M et al (2022) Relevant impact of surgery on quality of life for adolescent idiopathic scoliosis: a registry-based two-year follow-up cohort study[J]. Bone Joint J 104(2):265–273

19. Chapek M, Kessler A, Poon S et al (2023) The effect of adolescent idiopathic scoliosis on natural delivery and epidural use in pregnant females: a matched cohort study[J]. Spine 48(12):E188–E195
20. Hariharan AR, Shah SA, Petfield J et al (2022) Complications following surgical treatment of adolescent idiopathic scoliosis: a 10-year prospective follow-up study[J]. Spine Deformity 10(5):1097–1105
21. Kim G, El Sammak S, Michalopoulos GD et al (2022) Comparison of surgical interventions for the treatment of early-onset scoliosis: a systematic review and meta-analysis[J]. J Neurosurg Pediatr 31(4):342–357
22. Akazawa T, Kotani T, Sakuma T et al (2023) Health-related quality of life of patients with adolescent idiopathic scoliosis at least 40 years after surgery[J]. Spine 48(7):501–506
23. Willoughby KL, Ang SG, Thomason P et al (2022) Epidemiology of scoliosis in cerebral palsy: a population-based study at skeletal maturity[J]. J Paediatr Child Health 58(2):295–301

# Academic Performance in Blended Learning Environment Utilizing MOOCs



Anupriya Sharma Ghai, Ramesh Chander Sharma, Sanjay Jasola,  
and Alin Zamfirou

**Abstract** Blended learning is the combination of online and classroom instruction. It has changed the face of traditional education by allowing a custom-fit lesson plan. By integrating Massive Open Online Courses (MOOCs), students can watch lectures from subject matter experts wherever and whenever. The main focus of every type of education, whether it is online or traditional learning, examines the academic performances of the students. This paper provides an in-depth study on the evaluation of academic performance among undergraduate and postgraduate levels using MOOCs in a blended learning context. The results indicate differential effects on student performance by course type. It also reviews how methods of evaluation like quizzes, assignments, and exams measure the active involvement in student learning. This paper will help educators and institutions regarding the proper implementation of MOOCs in a blended learning context.

**Keywords** Blended learning · MOOCs · Academic performance

---

A. S. Ghai (✉)

School of Computing, Graphic Era Hill University, Dehradun, India

e-mail: [anupriya@gehu.ac.in](mailto:anupriya@gehu.ac.in)

R. C. Sharma · S. Jasola

Department of Computer Science and Engineering, Graphic Era Hill University, Dehradun, India  
e-mail: [rcsharma@gehu.ac.in](mailto:rcsharma@gehu.ac.in)

A. Zamfirou

Bucharest University of Economic Studies, Bucharest, Romania

National Institute for Research and Development in Informatics—ICI Bucharest, Bucharest,  
Romania

A. Zamfirou

e-mail: [alin.zamfirou@csie.ase.ro](mailto:alin.zamfirou@csie.ase.ro)

## 1 Introduction

Digitization has transformed almost every aspect of life, and education is no different. The use of digital technologies in educating and learning is pervasive and of great importance for educationists around the world who are interested in understanding how knowledge is acquired. Digital technology changes fundamental ways to impart as well as share this knowledge with others [1]. The same has pushed traditional teaching–learning procedures with a lot more components available. ICT and Digitization of Education ICT, on its part, ensures knowledge transfer from the teacher to the student. This has made educational resources more accessible; (i) Improved teaching and learning with interactive, digital textbooks or hybrid models, including between labs at different locations as well as classrooms, for collaborative learning leading to tailored experiences. (ii) Flexible, convenient working and enhanced assessment. Flexibility and convenience regarding the conditions in which one could work, e.g., independently at home or at any available time, as well as what he should do according to his speed. (iii) The method for assessing and providing feedback also improved. ICT and digitization hold immense promise in transforming the experience of education, providing opportunities to prepare students for a life organized around digital resources such as hardware devices, software programs, and the Internet. Moreover, digitization has provided a greater range of educational resources to students and educators (e.g., digital libraries and open educational resources). MOOCs provide access to a large amount of content and learning materials that can be used as complementary material in traditional classroom instructions [1]. Technology has allowed teachers to design flexible and stimulating classrooms that meet the diverse needs of students. Traditional education can be difficult to access—geographically, financially or otherwise—but online learning platforms and resources have made it possible for more individuals to get an education. With an internet connection, people can now learn and consume learning content anywhere in the world at their own pace and convenience [2].

The concept of e-learning embraces various forms of knowledge that are imparted online. Several forms of e-learning have emerged, including synchronous, asynchronous online learning, MOOCs, and blended learning. In all these various forms, it offers flexibility, personalized experience, interactivity, opportunities for collaboration, and catering to the diverse needs and preferences of learning. Which specific e-learning approach to adopt naturally depends on the learning objectives, audience, subject matter, and other available resources [3]. MOOCs have gained fame due to their high-quality educational content provided by leading institutions and experts; hence, they are open courses for all in any discipline, enabling learners from any part of the world to take courses at their speed and time. With MOOCs, education has become more accessible and more affordable. Learners can easily get more knowledge and improve their career opportunities all over the world [4]. Blended learning describes a situation where traditional face-to-face learning merges with online learning activities, creating a hybrid learning environment. It combines the

best of contact/traditional learning methods and the best of digital learning methods to afford an enhanced overall learning experience [5].

Blended learning incorporating MOOCs combines the advantages of online learning using MOOCs with face-to-face instruction within a blended learning setting. Although the integration of MOOCs in blended learning environments has begun to grow, the impact of blended learning that integrates MOOCs on the academic performance of students remains to be researched [3, 5]. Although blended learning with MOOCs is promising to promote engagement, flexibility, and access to educational resources, how such a learning modality influences academic achievement regarding the student's role is not clearly understood. Since there remains an empirical evidence gap and comprehensive analyses in the linkage between the use of MOOC-based blended learning and the academic performance of students, the necessity for a systematic review arises. It thus requires a critical review of how MOOC-based blended learning affects academic performance. Further, it informs practitioners in education, policymakers, and researchers about the efficacy and effectiveness of this approach on student learning outcomes. Various reasons for selecting MOOCs and different assessment practices to improve academic performance have been assessed in this present study.

## 2 Literature Review

Blended learning seamlessly integrates online learning activities and resources with face-to-face instruction. It combines in-person classroom interactions, such as lectures, discussions, and hands-on activities, with digital content, virtual conferences, multimedia resources, and online assessments. Here are definitions of blended learning provided by different researchers. Defined by Graham [6], blended learning is combining traditional face-to-face classroom teaching methods with online learning activities. Also, combine traditional and online learning to enhance the learning experience. Blended learning is an instructional approach that combines elements of online learning and face-to-face instruction. The focus is on integrating online and face-to-face elements within the instructional practice [7]. Blended learning is “the thoughtful integration of classroom face-to-face learning experiences with online learning experiences.” They emphasize the integration of both modalities to create a cohesive learning environment [8]. Picciano [9] states that blended learning involves strategically combining traditional classroom instruction with online learning activities. The author emphasizes the strategic approach of combining the two modes of education [9]. The above definitions provide a consistent understanding of blended learning as integrating traditional face-to-face instruction with online learning activities. The paper [10] explores blended flipped learning, which combines face-to-face teaching with online resources, to enhance student engagement and learning experiences. It discusses the blended approach's benefits, challenges, and best practices. The blended learning framework emphasizes the integration of face-to-face and online learning environments. The paper provides

guiding principles and practical guidelines for implementing blended learning in higher education [8].

### ***MOOCs in Education***

MOOC stands for Massive Open Online Course, which is designed to accommodate an enormously large number of participants, sometimes in the thousands or hundreds of thousands. Many educational institutions, universities, and online learning platforms offer MOOCs. The key features defining MOOCs are due to [11, 12]: (i) Massive, meaning these are designed to accommodate a high number of participants worldwide. The course materials, including lectures, readings, and assessments, are available to a broad audience. (ii) Open to anyone interested in the subject matter without strict prerequisites or admission requirements. They provide access to educational content to learners regardless of their location, background, or prior academic achievements. (iii) Online delivered through various platforms, allowing learners to access course materials and participate in activities remotely. This flexibility enables learners to engage with the course at their own pace and from their preferred location with an Internet connection. (iv) Course structure follows a curriculum with specific learning objectives. They may include video lectures, interactive quizzes, discussion forums, assignments, and assessments to facilitate learning and knowledge retention. The evolution of distance education pedagogy, including the emergence of MOOCs as online learning [13].

MOOCs' completion rates and factors contributing to attrition provide insights into the challenges and opportunities associated with MOOC-based education [14]. A systematic study of the literature on MOOCs published between 2008 and 2012, analyzing the trends, key themes, and research methodologies used in this area [15]. Learner engagement and disengagement in MOOCs, using data-driven analysis that pinpoints distinct subpopulations of learners and their respective behavioral patterns [16]. Student dropouts in MOOCs using predictive modeling techniques aimed at early warning systems to pinpoint at-risk learners [17]. MOOCs can be used in a blended learning context to extend some pedagogical methods and meet a wide range of educational needs. MOOCs can serve as a kind of resource bank for students in blended learning classrooms [18]. The incorporation of MOOCs into blended learning frameworks can significantly improve student motivation and engagement through the provision of self-paced learning opportunities, collaborative experiences, and access to a diverse array of content and expertise [19].

### ***Blended Learning Using MOOCs***

Blended learning using MOOCs had a profound impact on education. It offers customized learning paths, flexibility, and accessibility, allowing learners to engage with various online resources at their own pace. Integrating online and in-person interactions promotes active participation, collaboration, and critical thinking [20]. Blended learning using MOOCs has transformed the educational landscape and expanded learning possibilities for individuals worldwide [21]. Some authors emphasize that integrating MOOCs into blended learning can enhance learner engagement,

resource access, and flexibility while promoting collaborative and active education. MOOCs' roles and benefits in blended learning may vary depending on the instructional context and learner needs. The benefits of integrating online and face-to-face learning are that combining MOOCs with traditional instruction can lead to improved learning outcomes, increased engagement, and enhanced flexibility for learners [5]. Means and Toyama (2010) evaluated the meta-analysis of online learning studies. They found that blended learning, including MOOCs, resulted in better student outcomes than face-to-face or strictly online instruction [22]. The instructors and students reported positive experiences when incorporating MOOCs into blended learning settings, including improved access to diverse content, increased learner engagement, and opportunities for self-paced learning [23]. Several studies have investigated the relationship between students' unawareness of MOOCs and academic performance. A study with first-year university students. They found that students utterly unaware of MOOCs had lower academic performance than those with some knowledge or experience with MOOCs [24]. This suggests a lack of awareness about MOOCs may hinder students' academic progress.

Furthermore, the effects of MOOC participation on academic performance found that students who actively engaged with MOOCs during their first year demonstrated improved academic performance compared to those who did not, indicating that early exposure to MOOCs can positively impact students' learning outcomes. Students who were initially unaware of MOOCs faced challenges navigating online platforms, accessing course materials, and engaging with interactive features. This lack of digital literacy skills hindered their ability to benefit from MOOCs, potentially impacting their academic performance [26].

### 3 Limitations, Challenges, and the Gap in Existing Literature

The perspectives of faculty members who have implemented blended learning with MOOCs. The authors discuss challenges such as designing practical learning activities, managing large enrolments, and integrating MOOC content with face-to-face instruction [27]. The challenges of integrating MOOCs into blended learning models include concerns about quality assurance, faculty roles, student engagement, and scalability [28]. The MOOC research initiative data and identify challenges related to implementing blended learning with MOOCs, such as the need for effective learning analytics, strategies for learning engagement, and support for diverse learner populations [29]. A case study on the challenges of designing and implementing a blended MOOC model. The authors discuss challenges related to instructional design, faculty support, student engagement, and assessment in a blended learning context [30]. The challenges students and instructors face in using MOOCs within blended learning environments. It highlights challenges related to learner motivation, time management, self-regulation, and instructor workload. The potential of blended learning with

MOOCs to enhance academic performance. The positive effects of blended learning with MOOCs on academic performance are based on research conducted around that time [32]. The impact of blended learning using MOOCs on academic performance in different educational settings [33]. The integration of blended learning and MOOCs as credit courses was described at Graphic Era Hill University in Dehradun, India. Where the university has promoted blended learning using MOOCs since 2014 [34, 35]. The approach involved combining a traditional classroom setting with online courses, where students were advised to enroll in any 6- to 8-week MOOC of their choice from different platforms.

## 4 Research Methodology

This paper investigates a developed approach for delivering and encouraging a blended learning approach through the integration of MOOCs. Students were allowed to select a MOOC from existing platforms under three categories: (i) A topic based on their hobby or passion. (ii) A topic pertinent to their academic program that is not encompassed within the curriculum. (iii) A topic that aligns with their course and is part of the syllabus.

The course was designed as a single credit awarding course, requiring a course pass but an optional certificate presentation. This optional certification depended on potential fee requirements associated with some courses. The progress and learning course path in the MOOC were measured through formative assessments. Learners were required to present their progress through three different stages:

- Stage 1. Signup and orientation to course content.
- Stage 2. Midway through the course progress report.
- Stage 3. Final report and presentations upon course completion.

Marks were assigned at each stage to determine the end grade and the credit given. Professors thus guided students on course selection to help them gain more confidence in their field of study and future careers. However, the decision was upon the students, which could be driven through faculty or the students' peers.

### *Sample*

The research examined a sample of 200 students enrolled in the initial year cohorts of the undergraduate and postgraduate. The investigation took place during the academic period spanning July to December 2023.

### *Data Analysis*

The research would involve a comparison of the students' ending grades in the MOOC-taught course to their performance in the previous year. Furthermore, demographic variables are addressed to find out how they influence learning outcomes in a blended learning context.

### ***Research Questions***

RQ1. How does age influence students' academic performance in blended learning using MOOCs?

RQ2. Is there any significant difference in academic performance between male and female students in blended learning using MOOCs?

RQ3. How does the course selection relate to the academic performance of students through blended learning using MOOCs?

RQ4. To what level would students' academic performance be influenced through computer literacy within blended learning using MOOCs?

RQ5. How does the choice of MOOCs affect the student's academic achievement in blended teaching?

## **5 Data Analysis**

The analysis investigates factors in the academic performance of students included in a blended MOOC learning environment were analyzed in this study. Five main research questions were addressed; these relate to different aspects of student performance, such as the influence of age, gender, course selection, computer skills, and workload. By using sensitive number analyses, such as correlation coefficients, regression models, T-tests, ANOVA, and Tukey's HSD tests, the study brings to light some key trends and relationships that drive academic outcomes. These findings provide useful insights into how these factors interact and influence students' success or difficulties within blended learning environments, thus informing possible ways of enhancing teaching methods and supporting various types of students.

A detailed interpretation of the correlation coefficient in the context of Research Question 1. Table 1 presents the relationship between age group and academic performance, highlighting the percentages of positive and negative performance and the corresponding correlation coefficients.

In the table above, it is provided the average percentage performance for every age group, with respect to the correlation coefficient between age and academic performance. This is confirmed by the negative nature of most correlation values, which

**Table 1** Relationship between age group and academic performance

Age Group	Positive performance (%)	Negative performance (%)	Correlation
18-20	26.53	73.47	-0.327
20-22	22.92	77.08	-0.452
UPTO 18	15.79	84.21	-0.254
ABOVE 22	18.75	81.25	-0.132

**Table 2** Regression analysis between age and academic performance

Variable	Coefficient	Standard error	p-value
Intercept	12.85	4.21	< 0.001
Age	-1.15	0.53	0.037

further means a negative relationship, suggesting that as age increases, performance tends to decrease. The higher the correlation coefficient, the stronger the correlations. Also, the regression analysis was done to study how age relates to academic performance in blended learning with MOOCs. The regression model gave the following equation:

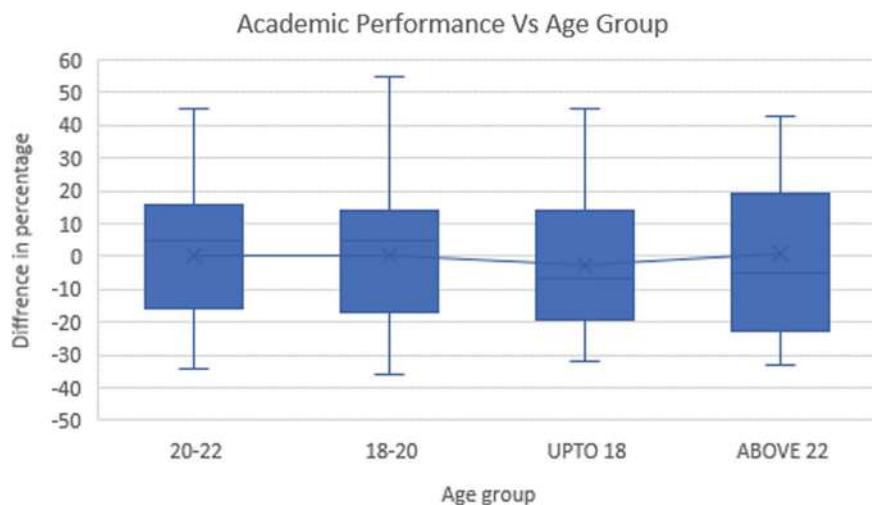
$$\text{Academic Performance} = -1.15 * \text{Age} + 12.85$$

Table 2 summarizes the results of the regression analysis, showing the coefficients, standard errors, and *p*-values for the intercept and age variables in relation to academic performance.

The coefficient for age is -1.15, indicative that for each one-unit increase in age, the associated academic performance would fall by 1.15%. Looking at the intercept of 12.85, when age takes a value of zero, the academic performance would be at a mean level of 12.85%. The test of significance of the variable age in the regression model is 0.037, less than 0.05. This infers that age is a significant determinant of academic performance in the context of a blended learning environment with MOOCs. The *R*-square value of 0.168 explains that about 16.8% of the variation in academic performance could be explained through the age variable in a regression analysis. That shows a good fit for the relationship between age and academic performance. From this, it is clear that age plays an important determining role in the academic performances of students in a blended learning environment with MOOCs. Figure 1 shows the students of older age usually perform lower compared to their peers.

Research Question 2, which investigates the difference in academic performance between male and female students in blended learning using MOOCs, the following table provides a detailed analysis. Table 3 outlines the average percentage performance and correlation between gender and academic performance, comparing male and female students in a blended learning environment using MOOCs.

In Table 3, the column “Gender” shows two groups of either male or female. The column “Average Percentage Performance” shows the performance of the male and female students separately in percentages. The “Correlation” column shows the correlation coefficient between gender and performance in school. According to the data, male students have an average performance of -8.6%, which is a little lower than female students, who have an average performance of -7.4%. The negative numbers show that both male and female students, on average, had a drop in their academic performance in blended learning with MOOCs. The correlation r values are -0.091 among males and -0.083 among females. Very weak negative relationships are reported between gender and academic performance. That is to say, associations between gender and academic performance are slightly weaker among male learners



**Fig. 1** Academic performance versus age group

**Table 3** Difference in academic performance between male and female students

Gender	Average percentage performance	Correlation
Male	-8.6%	-0.091
Female	-7.4%	-0.083

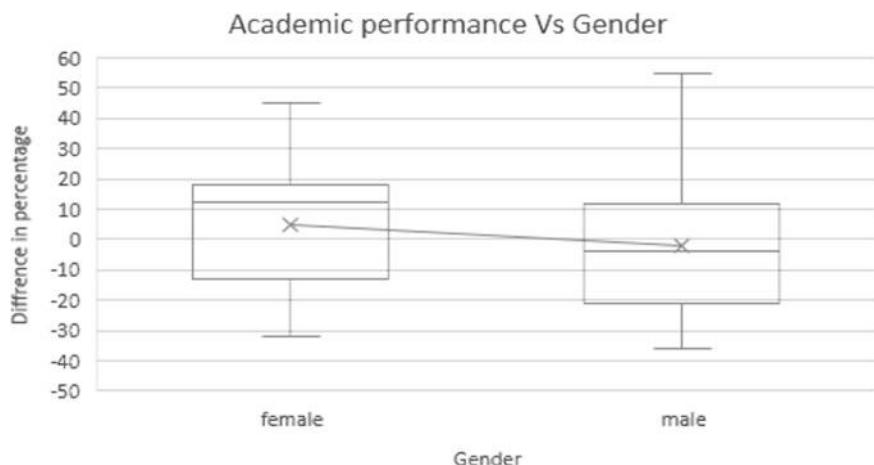
compared with females in blended learning with MOOCs. Taken together, the results of these analyses indicate that there is a somewhat different influence on the academic performance of male and female students, even though the association by gender is weak. That is, in blended learning with MOOCs, other factors could play a more important role in terms of academic performance than gender.

A T-test analysis was done to compare the average academic performance scores of male and female students. The results in Table 4 showed a significant difference between the two groups.

In the table above, the mean academic performance score for male students was 15.72, while the mean for female students was 17.98. This indicates in Fig. 2 that, on average, female students outperformed male students in blended learning using MOOCs. These findings suggest that gender may influence academic performance in the context of blended learning using MOOCs.

**Table 4** Results of the T-test

Gender	Mean score	Standard deviation
Male	15.72	10.65
Female	17.98	9.85



**Fig. 2** Academic performance versus gender

The analysis for Research Question 3 involved examining the differences in student satisfaction between the undergraduate and postgraduate courses—the statistical analysis employed two tests: ANOVA and Tukey's HSD. The ANOVA test was conducted to determine if there was a significant difference in satisfaction levels between the two courses. Table 5 shows the results of the ANOVA test indicated a *p*-value less than 0.05, which suggests a statistically significant difference in student satisfaction between the two courses.

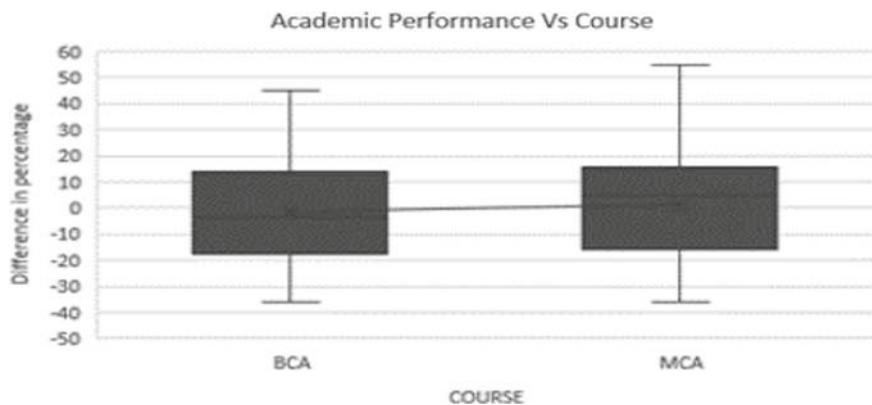
Table 5 shows the results of the ANOVA test, indicating that the model's *F*-value is 4.22, with a corresponding *p*-value of 0.042. Since the *p*-value is less than the significance level of 0.05, we can conclude that there is a statistically significant difference in student satisfaction between the two courses.

#### **Tukey's HSD Test**

A post hoc analysis using Tukey's HSD test was conducted to investigate further the nature of the difference in student satisfaction between the two courses. The means between the two groups, UG and PG, focus on the difference in their average scores. The "Difference in Means" shows that the undergraduate group's mean score is 7.31 points lower than the postgraduate group's mean. The confidence interval for this difference is provided by the "Lower Bound" and "Upper Bound" columns, which

**Table 5** Results of the ANOVA test

	The sum of squares (SS)	Degrees of freedom (df)	Mean square (MS)	<i>F</i> -value
Model	168.36	1	168.36	4.22
Error	1542.44	118	13.07	
Total	1710.80	119		



**Fig. 3** Academic performance versus course

are  $-13.89$  and  $-0.73$ , respectively. This interval suggests that, with a certain level of confidence, the true difference between the two groups likely falls between these two values, indicating that the undergraduate group consistently scores lower than the postgraduate group. In Fig. 3, results suggest that the postgraduate course, within the context of blended learning using MOOCs, has higher student satisfaction levels than the undergraduate course.

The results for Research Question 4: To what extent does computer literacy affect students' academic performance in blended learning using MOOCs? Table 6 shows the relationship between computer literacy and mean percentage performance.

The data in the above table illustrates the impact of computer literacy on mean percentage performance. Individuals with low computer literacy have a positive mean performance of  $7.1\%$ , accompanied by a standard deviation of  $14.2$ . In contrast, those with medium literacy experience a negative mean performance of  $-10.4\%$ , with a slightly higher standard deviation of  $15.4$ . Interestingly, individuals with high computer literacy show an improved mean performance of  $5.3\%$ , although their standard deviation is the highest at  $15.7$ . This suggests that while higher computer literacy generally correlates with better performance, there is greater variability in performance among those with higher literacy levels.

Table 7 presents an ANOVA summary, showing a significant difference between groups where SS = Sum of Squares, DF = Degrees of Freedom, MS = Mean Square, and  $F$  = F-statistics.

**Table 6** Descriptive statistics for computer literacy levels

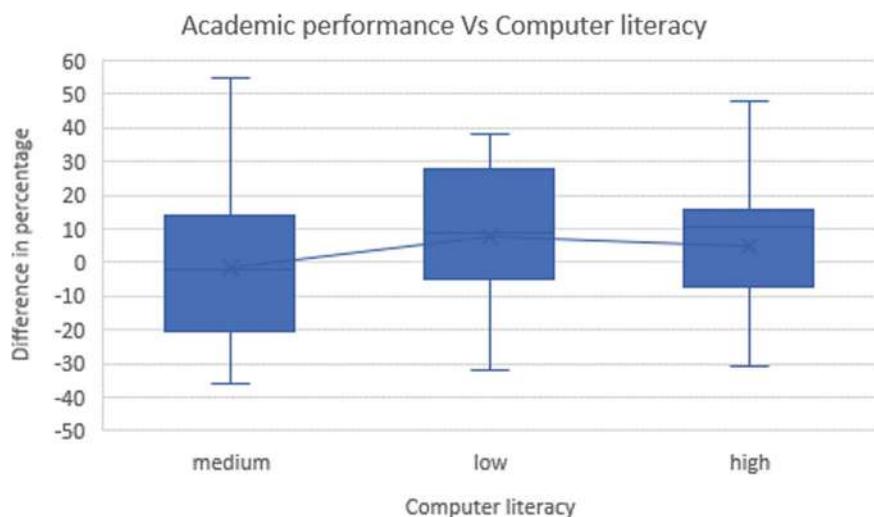
Computer literacy	Mean percentage performance	Standard deviation
Low	7.1	14.2
Medium	-10.4	15.4
High	5.3	15.7

**Table 7** ANOVA test results for computer literacy levels

Source	SS	DF	MS	F	p-value
Between groups	833	2	417	3.98	0.025
Within groups	18,078	195	92		
Total	18,911	197			

Table 7 illustrates the ANOVA test results indicate a significant effect of computer literacy on student's academic performance ( $F = 3.98, p < 0.05$ ). This suggests that the differences in academic performance among students with different computer literacy levels are not due to chance but somewhat attributable to the varying levels of computer literacy. Based on the mean percentage performance values, it can be observed that students with low computer literacy have a relatively higher mean performance compared to those with medium computer literacy. However, students with high computer literacy exhibit a slightly higher mean performance than low and medium computer literacy groups. These findings suggest that computer literacy plays a significant role in students' academic performance in blended learning using MOOCs. Figure 4 signifies that students with higher computer literacy levels tend to perform better. In contrast, those with lower computer literacy may face challenges in achieving academic success in the blended learning environment.

To properly analyze research question 5, statistical results for workload in both positive and negative academic performance:

**Fig. 4** Academic performance versus computer literacy

### ***Positive Academic Performance***

The mean workload (number of assignments and quizzes) for positively performing courses is approximately 7.3. This indicates that, on average, students with positive academic performance have a moderate workload. The standard deviation for workload in positively performing courses is around 2.8. This suggests that the workload varies among these courses, with some having a slightly higher or lower workload than the mean. The range of workload in positively performing courses is from 4 to 11. This indicates that the number of assignments and quizzes can vary between courses but generally falls within this range.

### ***Negative Academic Performance***

The mean workload for negatively performing courses is approximately 8.5. This suggests that, on average, students with negative academic performance have a slightly higher workload than those with positive performance. The standard deviation for workload in negatively performing courses is around 3.5. This indicates that the workload varies more among these courses than positively performing courses, with some having a significantly higher or lower workload than the mean. The range of workload in negatively performing courses is from 6 to 14. This indicates that the number of assignments and quizzes can vary widely between courses, with some courses having a relatively higher workload. These statistical results suggest that workload alone does not determine academic performance. While positively performing courses tend to have a slightly lower average workload, there is considerable variation within both positive and negative performance categories.

## **6 Results and Findings**

The study investigated several research questions related to students' academic performance in blended learning using MOOCs. The findings provide valuable insights into the influence of various factors on academic outcomes in this learning environment. In terms of age (RQ1), the analysis revealed that age does not significantly impact a student's academic performance. This suggests that students of different age groups can perform equally well in blended learning using MOOCs. Therefore, educators should focus on providing effective instructional strategies and support regardless of students' age.

Regarding gender (RQ2), the results indicated no significant difference in academic performance between male and female students. Both genders achieved similar average grades, emphasizing that gender is not a determining factor in academic success in this learning context. It is essential to promote equal opportunities and create an inclusive learning environment that caters to all students' needs, regardless of gender.

The relationship between the course and academic performance (RQ3) yielded significant findings. Specific courses were associated with higher academic performance, while others showed lower performance. This highlights the importance of course design, content, and instructional strategies in influencing students' educational outcomes. Educators should carefully consider the selection and development of courses to optimize student learning and achievement.

The influence of computer literacy on academic performance (RQ4) was found to be significant. Students with higher levels of computer literacy tended to achieve better educational outcomes. This underscores the importance of equipping students with the necessary digital skills and providing adequate training and support to enhance their engagement and success in blended learning using MOOCs.

The impact of MOOC selection on academic performance (RQ5) was also significant. Certain MOOCs were associated with higher academic performance, while others showed lower performance. Educators should carefully evaluate and select MOOCs that align with the learning objectives and preferences of the students to enhance their academic performance. A thoughtful and strategic approach to MOOC selection can positively impact student engagement and achievement.

In conclusion, age and gender were found to have no significant influence on academic performance in blended learning using MOOCs. However, course selection, computer literacy, and the choice of MOOCs were identified as influential factors. Educators should focus on designing effective courses, promoting computer literacy among students, and selecting appropriate MOOCs to optimize students' academic performance in blended learning using MOOCs. By considering these factors, educators can create a conducive learning environment that fosters student success and engagement in blended learning using MOOCs.

## References

1. Welsh DH, Dragusin M (2013) The new generation of massive open online course (MOOCs) and entrepreneurship education. *Small Bus Inst J* 9(1):51–65
2. Rodriguez CO (2012) MOOCs and the AI-Stanford like courses: two successful and distinct course formats for massive open online courses. *Eur J Open, Distance E-Learning*
3. Sahasrabudhe V, Kanungo S (2014) Appropriate media choice for e-learning effectiveness: role of learning domain and learning style. *Comput Educ* 76:237–249
4. Pilli O, Admiraal W, Salli A (2018) MOOCs: Innovation or stagnation? *Turk Online J Distance Educ* 19(3):169–181
5. Singh J, Steele K, Singh L (2021) Combining the best of online and face-to-face learning: hybrid and blended learning approach for COVID-19, post vaccine, & post-pandemic world. *J Educ Technol Syst* 50(2):140–171
6. Graham CR (2006) Blended learning systems. In: Bonk CJ, Graham CR (eds) *The handbook of blended learning: global perspectives, local designs*. Pfeiffer Publishing, San Francisco, pp 3–21
7. Vaughan N (2007) Perspectives on blended learning in higher education. *Int J E-Learn* 6(1):81–94
8. Garrison DR, Vaughan ND (2008) *Blended learning in higher education: framework, principles, and guidelines*. John Wiley, San Francisco

9. Picciano A (2009) Blending with purpose: the multimodal model. *J Res Cent Educ Technol* 5(1):4–14
10. Thai NTT, De Wever B, Valcke M (2017) The impact of a flipped classroom design on learning performance in higher education: looking for the best “blend” of lectures and guiding questions with feedback. *Comput Educ* 107:113–126
11. Pilli O, Admiraal W (2016) A taxonomy of massive open online courses. *Contemp Educ Technol* 7(3):223–240
12. Anders A (2015) Theories and applications of massive online open courses (MOOCs): the case for hybrid design. *Int Rev Res Open Distrib Learn* 16(6)
13. Anderson T, Dron J (2011) Three generations of distance education pedagogy. *Int Rev Res Open Distrib Learn* 12(3):80–97
14. Jordan K (2015) Massive open online course completion rates revisited: assessment, length and attrition. *Int Rev Res Open Distrib Learn* 16(3):341–358
15. Liyanagunawardena TR, Adams AA, Williams SA (2013) MOOCs: a systematic study of the published literature 2008–2012. *Int Rev Res Open Distrib Learn* 14(3):202–227
16. Kizilcec RF, Piech C, Schneider E (2013) Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In: Proceedings of the third international conference on learning analytics and knowledge pp 170–179
17. Yang D, Sinha T, Adamson D, Rosé CP (2013) Turn on, tune in, drop out: anticipating student dropouts in massive open online courses. In: Proceedings of the 2013 NIPS Data-driven education workshop vol 11. p 14
18. Bozkurt A, Akgün-Özbek E, Zawacki-Richter O (2017) Trends and patterns in massive open online courses: Review and content analysis of research on MOOCs (2008–2015). *Int Rev Res Open Distrib Learn* 18(5):118–147
19. Hew KF, Cheung WS (2014) Using blended learning: evidence-based practices, vol 20. Springer, Singapore
20. Archambault L, Leary H, Rice K (2022) Pillars of online pedagogy: a framework for teaching in online learning environments. *Educ Psychol* 57(3):178–191
21. Ali W (2018) Transforming higher education landscape with hybrid/blended approach as an evolving paradigm. *J Adv Soc Sci Humanities* 3(7):143–169
22. Means B, Toyama Y, Murphy R, Bakia M, Jones K (2010) Evaluation of evidence-based practices in online learning: a meta-analysis and review of online learning studies. Center for Technology in Learning, Menlo Park, CA
23. Bolliger DU, Martin F (2018) Instructor and student perceptions of online student engagement strategies. *Distance Educ* 39(4):568–583
24. Smith K, Jagesic S, Wyatt J, Ewing M (2018) AP® STEM participation and postsecondary STEM outcomes: focus on underrepresented minority, first-generation, and female students. College Board
25. Brown M (2021) What are the main trends in online learning? A helicopter view of possible futures. *Asian J Distance Educ* 16(2)
26. Chen T, Peng L, Yin X, Rong J, Yang J, Cong G (2020). Analysis of user satisfaction with online education platforms in China during the COVID-19 pandemic. *Healthcare MDPI* vol 8(3). p 200
27. Conrad D, Openo J (2018) Assessment strategies for online learning: engagement and authenticity. Athabasca University Press
28. Yuan L, Powell SJ (2013) MOOCs and open education: implications for higher education. *Int Rev Res Open Distrib Learn* 15(5):134–176
29. Gašević D, Kovancić V, Joksimović S, Siemens G (2014) Where is research on massive open online courses headed? A data analysis of the MOOC Research Initiative. *Int Rev Res Open Distrib Learn* 15(5):134–176
30. Hachey AC, Conway KM, Wladis C, Karim S (2022) Post-secondary online learning in the US: an integrative review of the literature on undergraduate student characteristics. *J Comput High Educ* 34(3):708–768

31. Lu OH, Huang AY, Huang JC, Lin AJ, Ogata H, Yang SJ (2018) Applying learning analytics for the early prediction of students academic performance in blended learning. *J Educ Technol Soc* 21(2):220–232
32. Thrun S (2013) MOOCs: the future is here. *J Nurs Educ* 52(1):3
33. Islam MK, Sarker MFH, Islam MS (2022) Promoting student-centred blended learning in higher education: a model. *E-Learning and Digital Media* 19(1):36–54
34. Ahmad I, Jasola S (2017) Supplementing higher education with MOOCs: a case study. In: 2017 International conference on emerging trends in computing and communication technologies (ICETCCT), IEEE, pp 1–5
35. Anupriya Bisht RK, Gahtori P, Jasola S, Ghai K (2019) Toward acceptance of MOOCs in higher education: a perspective of Indian students. *Int J Innov Technol Explor Eng (IJITEE)* 8(10S2)
36. Garrison DR, Anderson T, Archer W (2000) Critical inquiry in a text-based environment: computer conferencing in higher education. *Internet High Educ* 2(2–3):87–105

# Efficient Model Used for IoT-Based Digital Forensics Using Blockchain



Esha Tripathi , Upendra Kumar, Surya Prakash Tripathi, and Abhay Kumar Tripathi

**Abstract** The field of digital forensic in the Internet of Things (IoT) has seen limited research in recent times. The expanding and unpredictable nature of IoT environments poses challenges for existing forensic tools, methods, and investigative models. This situation creates difficulties for law enforcement and forensic experts. To address these types of issues, this methodology proposes a blockchain dependent digital forensics solution for IoT surroundings. In this suggested framework, all IoT system communications are kept in form of blockchain transactions and this facilitating the formation of a robust and straightforward chain of custody mechanism. The usage of blockchain technique confirms the validity of data for analysis, enhances security, and improves the trustworthiness of integrity preservation through a decentralized approach. Additionally, the availability of public distributed records allows forensic investigation team members, including various device users, service providers, manufacturers, and investigators, to conduct transparent inquiries. To validate the concept, an analysis of the proposed framework was performed.

**Keywords** Distributed framework · Digital forensics · Blockchain · Internet of Things

---

E. Tripathi ( ) · U. Kumar

Institute of Engineering and Technology, Dr. A.P.J Abdul Kalam Technical University, Lucknow, India

e-mail: [tripathi.esha@gmail.com](mailto:tripathi.esha@gmail.com)

U. Kumar

e-mail: [ukumar@ietlucknow.ac.in](mailto:ukumar@ietlucknow.ac.in)

E. Tripathi · A. K. Tripathi

Pranveer Singh Institute of Technology, Kanpur, India

S. P. Tripathi

R R Institute of Modern Technology, Dr. A.P.J Abdul Kalam Technical University, Lucknow, India

## 1 1. Introduction

As Internet penetration has expanded and gadgets have shrunk in size, society has been more digitalized, electronics have grown increasingly widespread, and electronic gadgets are more efficient have been developed [1]. These accomplishments have resulted in the development of the IoT too, an ecosystem in which numerous gadgets communicate with one another. The complicated nature of the Internet of Things (IoT) framework, as well as the deficiency of an integrated norm, impede digital monitoring and make forensic evidence collection tough for safety and law enforcement groups [2]. Nevertheless, various national security organizations and agencies recognize that developing appropriate IoT survey criteria as well as solid security measures may assure the success of their investigations. Furthermore, present forensic techniques and communication ideals are incompetent of dealing with the IoT's and dispersed framework's extremely diverse nature [3]. The tremendous expansion of IoT has increased the risk of criminality, which might result in unanticipated consequences for most cyber criminals. Sensor sensitive data in complex physical systems might be susceptible, though it passes through another networks, communication pathways and user devices [4].

However, in 2009, Satoshi Nakamoto suggested novel technology that is blockchain [5]. Individual type of transactions are now safe and transparent, and previous centralized techniques are no longer required. Originally, consumer transactions were saved and supervised by a central organization, i.e., bank. Though, this centralized system can't be protected against malevolent attackers. With the help of sharing a single distributed ledger for every consumer, cryptocurrency based on blockchain may be exchanged without a centralized organization, i.e., bank [14]. This is due to the fact that all transactions require the approval of more than 50% of all consumers. Every 10 min, the integrity of each record is assured by demonstrating data based on transaction using Proof of Work (POW). Currently, verification methods used for data integrity in digital forensics are basically designed to allow investigators to acquire digital proof in line with legally procedure and image the disc using expert digital forensic equipment [12].

During this approach, an administrative body verifies digital evidence. Although, this centralized form of integrity preservation raises possibility of sign being altered with by hostile attackers [14]. As a result, this article presents a technique of integrity preservation based on ledger transparency that blockchain is used to digital forensic inquiry procedure [24]. The primary legacy of this article is following given as:

- This research elaborates issues and limits of digitalized forensics in IoT ecosystem.
- This presents the model for cyber forensics in Internet of Things context based on blockchain technique.
- To establish evidence of this concept, analyzed the suggested model and explored future study about research for directions.

This article examines blockchain, an existing relevant digital forensics investigation method, digital forensics in IoT in Sect. 2. Next, we present overall architecture, block structure, and workflow for the IoT environment in Sect. 3. Finally, Sect. 5 finish study though conclusion.

## 2 Literature Review

Pre-existing finance related transactions for all consumers controlled through centralized authority. The record is the sole method to ensure integrity of financial information of clients throughout this procedure. However, in the absence of a centralized authority such as a bank, bitcoin uses the blockchain mechanism to maintain integrity. Blockchain technology began as a rudimentary cryptocurrency system. When using the blockchain, consumer transactions not require a central authority. To preserve ledger transparency, ledgers of every cryptocurrency contestants are in shared mode and controlled through another contestants. Other contestants guarantee that contestants transactions are transparent. Due to the decentralized nature of information storage, hostile attackers or insiders cannot perform deception of information or manipulation assaults. Moreover, all existing transactions are completed with the permission of another contestants, preventing from malicious parties from the repeating or denying the transactions [4]. Everyone in the room is a manager or a supervisor. Based on the grade of transparency of ledger, a blockchain is characterized as public and private blockchains. Though, on the public blockchain, everyone shares and manages their ledgers. Though one version of blockchain, every contestants are subjects who basically manage and then validate ledgers. Contestants in public blockchain can record first then find any type of ledgers, and it share the common distributed ledgers, and then get incentives when it block in order to retain the network functioning. Though, because there is no central authority in this form of blockchain, this is incredibly difficult to modify or improve the blockchain framework. It is also inefficient due to the lengthy verification procedure. The essential authority controls the ledgers of all private blockchain contestants. Contestant in private type blockchain network is only probable if authorization is granted by the central authority. So this form of blockchain has an essential authority, it is simple to alter or improve the blockchain system [18]. However, consortium blockchain is a cross between private and public blockchains. Transactions occur in this kind of blockchain are carried out between already permitted parties. So, these three forms of blockchains can be employed based on the application of blockchain method. As public blockchain may be utilized for cryptocurrency transactions like Ethereum or Bitcoin, while in private blockchain, it can be used when an essential authority controls numerous contestants, and in consortium type blockchains, it can be utilized comparatively for minor transactions among trusted partners.

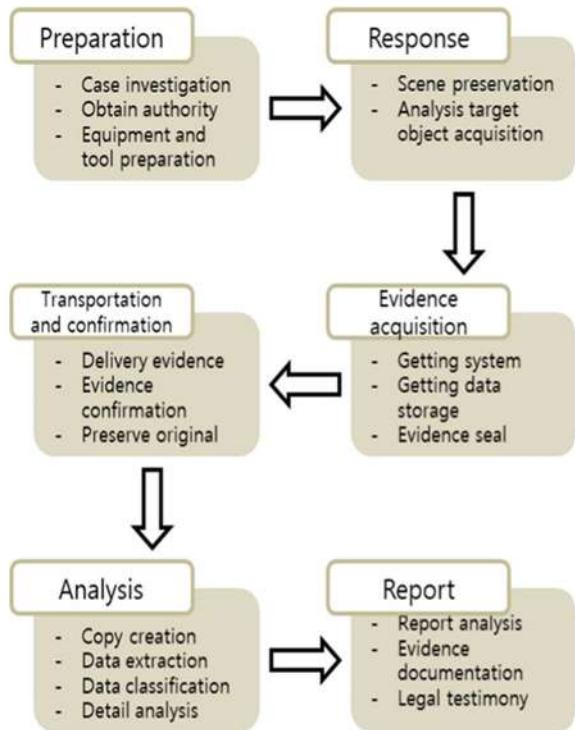
## 2.1 Pre-Existing Method for Digital Forensic

Digital forensic enquiry used for modern conventional systems, i.e., personal computers and various servers that contains of six phases: preparation, reaction, evidence collecting, transfer, confirmation, analysis, and make report. During investigation's preparation phase, preliminary effort is completed, like checking an incident and gaining authorization to validate it. During response phase, crime scene is maintained, and media to be studied is attained. During evidence acquisition process, the system, storage media are collected, and the evidence is wrapped. During the transmission and confirm stages, the wrapped proof is delivered to the investigating team, but original one is maintained. During investigation process, copy of data is made, data is selected, classified, and extensively examined. Finally, at report stage, analytical reports are produced and testified in the court [6]. Electronic forensic can also be predicted through intensity-level multi-fractal dimension feature with twin support vector machine [19, 25].

This procedure delivers data to an analyst, then evaluates data in compliance with the case manager's analysis need. The data integrity might be jeopardized if data source is analyzed at this time. As a result, data is cloned and imaged, allowing for the same analysis as the initial information. A write block-based device is utilized to keep actual data intact, and replication is done two times to discriminate between storage and analysis. It secures data by obtaining hashes of the original, analytical, and archive data [6]. The integrity of digital imaging activities is also protected, just as it is with data replication. As a result, Korea's present method for getting data and maintaining integrity in an electronic forensic inquiry is based on a central authority, like a gentle person, analytical specialist, Supreme Prosecutor's Office, or National Intelligence Service (NIS). If responsible person or expert is malevolent, the original data may be updated, putting their stability at risk, because the hashing method used to preserve integrity is only conducted once across the data.

The current digital research approach is inadequate to handle the heterogeneity and dispersion features of IoT-based infrastructures. As a result, unlike present forensic methodology and model, an IoT-based digital forensic findings necessitates the use of a technique designed specifically for IoT infrastructure. There are several kinds of digital mark at the place of crime scenes. Evidence may be located on external hard drive, suspect's PC, laptop, mobile devices (such as a smartphone or tablet), and USB. Furthermore, because digital media and storage devices have different capacities, digital forensic analysis of each device can require a significant amount of time and resources [7]. Further, for overcoming these challenges, electronic forensic study has been conducted for a lengthy duration. However, digital forensics must be arranged to deal with Internet of Things (IoT). IoT atmosphere is growing at a different speed than the present electronic ecosystem. Therefore, as a result, present digital forensic investigation technique has to be updated to match IoT framework. Figure 1 displays current digital forensic analysis procedure in real-world cyber-crime scenario. With the help of blockchain technology in IoT forensic investigations protects data integrity while also streamlining chain of custody processes.

**Fig. 1** Pre-existing procedure for digital forensic [10]



## 2.2 Digital Forensic in the Field of IOT

IoT forensics is the forensic study of different IoT-dependent frameworks which can be analyzed electronically with the help of traditional methodologies. It is classified into three main groups: cloud forensics, network forensics, and device forensics [3].

- Device forensics: Electronic evidence collecting on various systems in IoT scenarios. This forensic approach extracts digital evidence from physical devices including video, audio, near field communication (NFC), memory, and other Internet of Things devices [13].
- Cloud forensics: All types of IoT-based devices are connected through network and then share various resources in a virtual-based environment, allowing them for communication across network via cloud services. Major types of cybercrime in the cloud-dependent IoT scenario focus on cloud-created content. This happens due to complexity of systems increases which link to cloud, along with appearance of information, which is cloud-based and centralized.
- Network forensics: Digital forensic study basically target networks takes place out in IoT surroundings employing several forms of networks. This basically used to collect anomalous attack logs and then carry out forensic investigations. So,

this notion may be used to a household or a local area network (LAN), a wide area network (WAN), a metropolitan area network (MAN), and so on.

### **2.3 Existing Work Related to IOT-Based Digital Forensic**

This part specifies a summary of current digital forensics techniques and research in IoT-based ecosystem.

- Kebande and Ray [4] proposed a framework for digital forensic inquiry in the Internet of Things. This system consists of three parts: proactive processes, reactive procedures, and IoT forensics. Proactive approach is a digital forensic preparation methodology used in IoT forensics activities. Though, IoT-based forensic procedure covers a range of forensic scenarios for extracting evidence in an IoT context. The reactive method determines the digital forensic investigation procedure when a suspected security event is detected. With the help of this paradigm, new digital forensic investigation methodologies might be created to meet the demands of IoT's increasingly diverse and dispersed architecture.
- Zhang et al. [8] developed an approach to improve the dependability of the chain of custody in cloud-based forensic investigations. The digital forensic approach necessitates comprehensive documentation to establish chain of the custody. And data obtained during a forensic methodology reveals the source system in this investigation, which strengthens the chain of custody.
- Cebe et al. [10] developed a lightweight integrated block-based forensic system for smart car data analysis. The purpose of forensic approach is meeting legal concerns while objective is to demonstrate faulty vehicle in scenario of a traffic incident. Although, proliferation of the smart automobiles shows a new issues for the digital forensics in IOT systems. Whereas traditional forensic methods, the decision-making entity's sensor data may be leveraged to create an effective smart car digital forensic investigation strategy. There are several parties involved, including the manufacturer, driver, insurance company, and investigator. They suggested and explored a reliable smart car forensic architecture.
- Oriwoh et al. [11] presented a novel technique to IoT digital forensics. They emphasize that the Internet of Things is designed to be network having self-managing, decision-making models. Though the influence of IoT on forensics is huge in form of crime perpetrated through smart devices. So, effects of IoT interaction between several devices have a substantial influence on current digital forensics practices. They anticipate IoT digital forensics to differ from traditional digital forensics.
- MacDermott et al. [9] highlighted digital forensic limitations and problems in the IoT context. In their research, they differentiated weight of the electronic evidence from databases in a smart city atmosphere to address future of digital forensics in IoT scenarios. They also underlined the significance of a contemporary approach to digital forensic inquiry as the Internet of Things (IoT) age transitions to the Internet of Anything (IoA) era.

## 2.4 Problem Statement

Today, the majority of evidence preservation solutions rely upon centralized repository system comprised of third parties. As it is obviously causes a host complication. Certain safety criteria must always be met in centralized structures. Intrusion into a central storage node causes major issues such as data leaking and manipulation. In terms of transparency and dependability, the centralized system, like earlier digital forensic technologies, is vulnerable. Individuals are always skeptical about the reliability of services. Further, because various companies manufacture different IoT devices, a comprehensive digital forensic framework is required. This negatively affects forensic investigation and system scalability [16].

Distributed blockchain networks, on the contrary, offer a transparent and predictable secure environment in which content can be defended with the help of huge quantities of computing power. Trusted timestamps can be applied to freshly produced blocks instantly. Most crucially, trust difficulties may be avoided by dispersing the auditor's power. It possesses accuracy, timeliness, and integrity necessary for conservation.

## 2.5 Requirements for IOT Forensic

An Internet of Things (IoT) digital forensic examination must incorporate several key elements [10]:

- Integrity: The primary objective of digital forensic investigations is to gather evidence for legal proceedings. Maintaining honesty throughout the entire forensic process is crucial.
- Accountability: Experts should require liable for trial's outcome by offering unbiased evidence of content source and authenticity.
- Decentralization: The investigation method for forensic in IoT ecosystem should minimize support on a single object while ensuring reliability of all entities in form of distributed fashion.
- Continuous forensic investigation: Forensic experts should have manipulated to historical content even a cybercrime happens previously. A comprehensive framework for crime scene evaluation should established.
- Low overhead: Due to the high volume of interactions handled by IoT devices, they require minimal overhead at the endpoints (device layer).
- Data protection: Given the blockchain's nature of sharing all ledgers among participants, it is essential to prevent privacy breaches at all costs. Privacy violations must be strictly avoided due to the shared nature of blockchain ledgers among members.

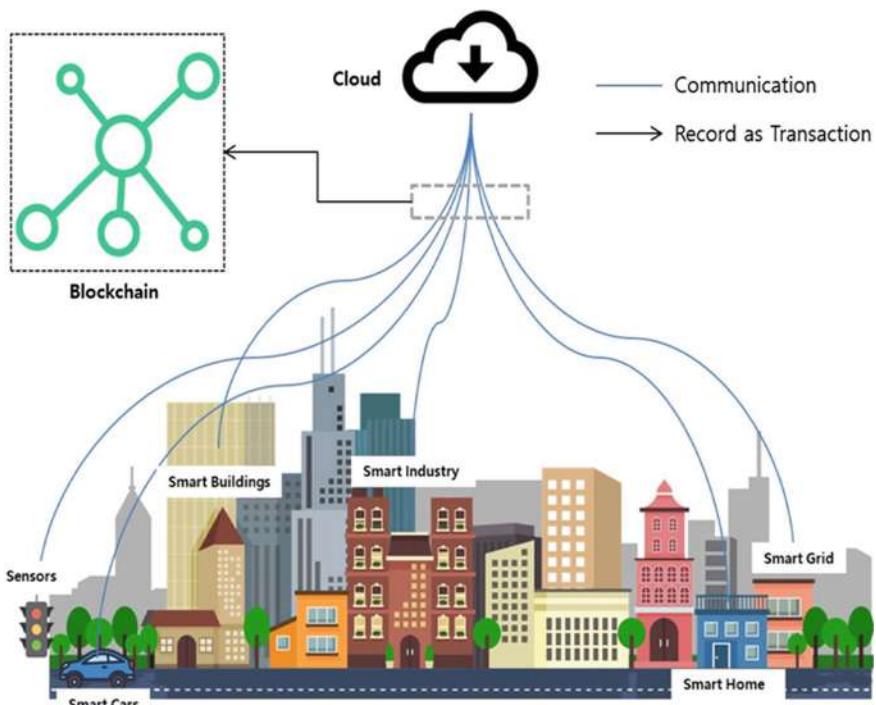
### 3 A Blockchain Oriented Digital Image Forensic Framework

This section offers a summary of the proposed model, including structure containing blocks inside it, blockchain participants, and procedure.

### 4 Summary of Suggested Framework

Figure 2 defines the suggested IoT digital forensics model. The suggested system comprises of three layers: cloud, blockchain, and IoT devices. Gadgets in an Internet of Things system regularly connect with the cloud. By 2020, the amount of Internet of Things devices is estimated to reach 26 billion [17]. Using traditional digital forensic processes to evaluate a large number of IoT devices is almost difficult in this case.

Sensors, smart autos, smart buildings, smart industries, smart homes, and smart grids are examples of IoT ecosystems. Cybercrime can occur in any of these sectors at any moment, and a robust forensic framework must be built to deal with it [15].



**Fig. 2** Summary of the suggested framework

IoT devices come in a variety of roles, services, manufacturers, technologies, and data formats.

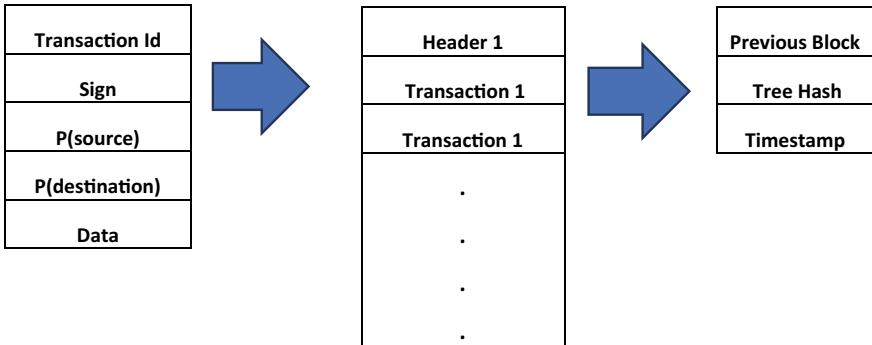
At the user's choosing, IoT devices transmit and receive vast volumes of data. Using the present forensic technique on every device with a high number of connections complicates the inquiry in this scenario. As an outcome, in the proposed architecture, data created during each IoT device's communication process is stored in the blockchain as a transaction. The digital forensic inspector utilizes block integrity storage and a streamlined chain of custody mechanism.

#### ***4.1 Block Structure of Suggested Framework***

This research proposes an alternative block structure to the one currently in use. The blockchain of the suggested framework employs blocks with a structure illustrated in Fig. 3. Each block is divided into two main components: the header and the transaction. The header is further segmented into blocks, including the Merkle tree hash with timestamp. A sequential identification number is assigned to each newly created block. The Merkle tree hash is utilized by investigators and other parties to identify blockchain transactions. The timestamp indicates the block's creation time. The transaction section consists of five elements: Transaction Id, digital signature, P (source), P (destination), and contents. The Transaction Id stores the SHA256 hash results of the signature, P (source), P (destination), and data, serving as the transaction's unique identifier. The digital signature is generated using the P (ID) and the private key of the sender's IoT device. P (source) records the P (ID) of the data-providing device, while P (destination) stores the P (ID) of the receiving device. The data section contains information generated through device communication. This study's proposed block structure differs slightly from the current one, utilizing blockchain not as a FinTech component but to ensure integrity in forensic investigations. Consequently, a block is conceptualized as a secure data storage mechanism between devices rather than an item for competitive mining.

The blockchain in intended digital forensics model is classified into variety of groups: IoT device user, investigator, service provider, and IoT device manufacturer. Blockchain participation in the present research are a modified form of Cebe et al. [10] suggested blockchain participant structure.

For enquiry of digital forensic, blockchain is utilized to ensure dependability, integrity, and accessibility of stored data, as well as to offer all stakeholders with a shared ledger for secured and efficient information handling. Unless everyone involved agree, the information on the blockchain cannot be changed. With the permission of government authorities, the crime scene investigator may have access to the IoT forensic blockchain. Participants apart from researcher can now access the blockchain information. This is why all blockchain participants exchange the same information in order to maintain the integrity of data. If an IoT device has been utilized to commit a crime, the user may utilize blockchain data to help in the investigation or to prove his innocence. Blockchain information might be used



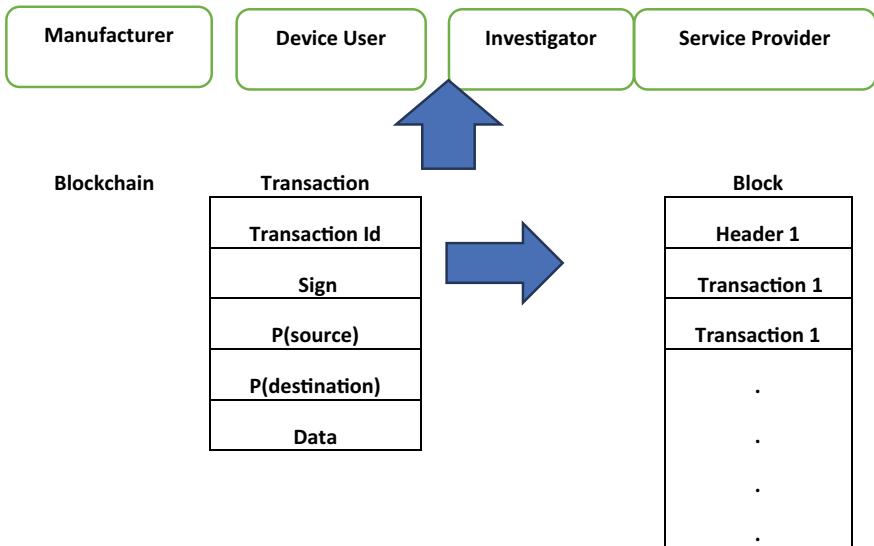
**Fig. 3** Block diagram for suggested framework

by IoT device makers to reveal faults in produced IoT items or to ensure quality. In the instance of a service provider, blockchain data might be utilized to ensure both client service and personal information. This section discusses privacy problems in forensic investigation and blockchain technology. If an IoT device is used to commit a crime, the user may utilize blockchain data to help in the investigation or to prove his innocence. Blockchain data might be used by IoT device producers to reveal faults in produced IoT items or to ensure quality. In instance of a service provider, blockchain data might be utilized to ensure both client service and personal information. This section discusses privacy problems in forensic investigation and blockchain technology.

## 4.2 Process for Suggested Framework

Figure 4 depicts the working of the proposed blockchain-based digital forensic model for the IoT domain. The suggested architecture is composed of three layers: the device layer (bottom layer), the blockchain layer (middle layer), and the participants' layer (top layer).

This approach uses two IoT devices as examples. At the device layer, every Internet of Things device talks and shares data. Each device has a key pair and a P (ID) for digital signatures, and it may be categorized as an IoT device. Using the data generated during each IoT device's connectivity, we build blocks in the blockchain layer, which is the intermediate layer. The data communicated from Device 1 to Device 2 is recorded in the data section at the bottom of the transaction. The P (ID) of data collected from sender Device 1 is saved in the P (source) of the transaction, but the PUF ID of data receiver Device #2 is kept in the P (destination). The transaction is then digitally signed with the sender Device 1's private key and P (ID). If the digital signature, P (source), P (destination), and data are all collected, the result is hashed twice with the SHA256 hash method and added to the transaction's transaction id.



**Fig. 4** Flowchart of suggested IoT-based framework

When the previous technique is finished, one transaction is performed, which is then continually logged in preparation for the next communication. If the number of transactions exceeds the block size, a new block is generated and connected to the preceding one. When a crime happens in the higher layer of the participants' layer, each participant (device user, manufacturer, service provider, and investigator) has the ability to view the blockchain's public ledger. For example, the investigator can validate that Device 1 is a valid sender by decrypting the transaction's digital signature using Device 1's public key. Each participant in this workflow may verify the integrity of the data received by the IoT device. Investigators, in particular, may lower the resources required by chain of custody standards designed to preserve data integrity and openness.

## 5 Experimental Analysis

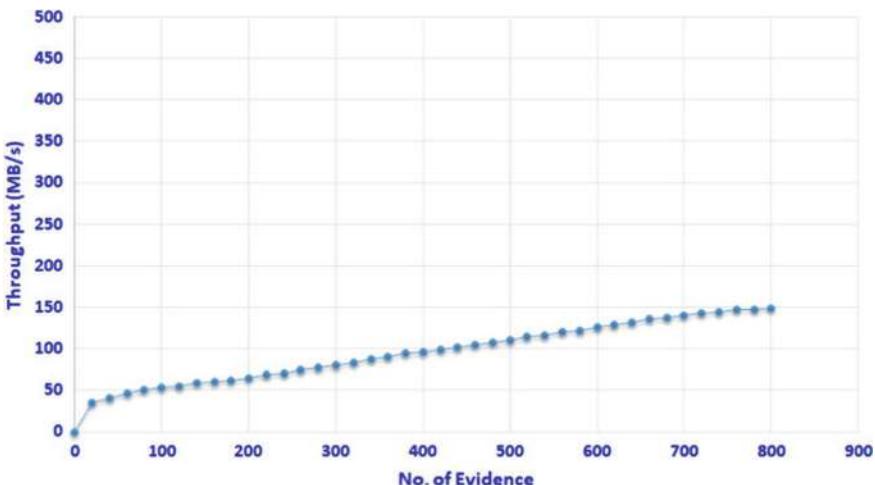
### 5.1 Experiment Environment

To validate the concept, we run the suggested model prototype on the Ethereum private network platform. With turing completeness and strong scalability, the Ethereum blockchain architecture is primarily meant for smart contracts [20]. We discuss Bitcoin as a blockchain approach in the introduction, but we chose Ethereum as an experimental platform for reasons such as ease of testing and openness of results. Ethereum, like Bitcoin, relies on the Proof of Work (PoW) consensus method. The

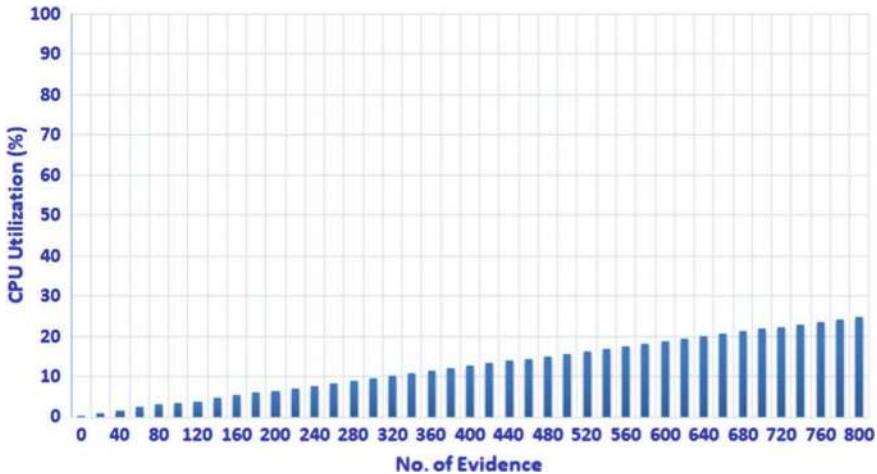
notion of a smart contract is vital in this study as the blockchain is utilized for documenting the transmission process of the IoT device as a transaction rather than for FinTech [23]. We created a private blockchain using Geth [21] and it tailored blockchain consensus and smart contracts for simulation. Geth is a command-line interface that permits you execute the whole Ethereum protocol. We used Mist Browser to develop smart contract interfaces for evidence production, acquisition, and report preparation [22]. Mist Browser is a useful tool for engaging with blockchain components like as Ether, smart contracts, and transaction blocks. Therefore, trial desktop included 64 GB DDR3 RAM and an Intel i7 processor. According to limited resources provided by laboratory, we are currently creating a prototype to examine the feasibility status of our suggested technique. We will improve the system model in the future to produce a wide-ranging forensic model.

## 5.2 Analysis

Moreover the data depicted in Fig. 4, we looked at gas related consumption in relation to blocksize and transaction count. The amount of gas specifies the cost paid by suggested model in generating the proofs, while size (bytes) denotes number of transaction blocks in our private blockchain network. As illustrated in Fig. 5, the transaction's evidence creation was simulated up to 800 times. The justification for limitation on number of proofs to 800 is because when simulating more than 800 evidences, the experimental desktop's performance is projected to have an effect on the experimental results.



**Fig. 5** Number of evidence with respect to throughput



**Fig. 6** Number of evidence with respect to utilization of CPU

In this simulation, the evidences produced execution time. We are presently working on a prototype to determine the practicality of our suggested technique. In the system model, we will increase the number of evidence generators. The connection between throughput and evidence creation is seen in Fig. 6. The throughput grows according to the number of participating nodes. As a result, one additional node may be added to the model, resulting in increased throughput. In addition, to address performance problems, we recoded the overall average CPU consumption vs the quantity of evidence produced. Figure 5 depicts how CPU consumption fluctuates according to the quantity of evidences provided. The suggested framework's scalability is demonstrated by a linear increase in CPU utilization as the amount of evidence generated grows.

## 6 Conclusion

Presently, digital forensics investigation is being conducted autonomously through essential authorities like the Police, Prosecutor's Office and National Intelligence Service. The efficiency and procedural convenience are appropriate, but if a hostile adversary targets the central authority, the integrity of possible evidence may be jeopardized. Furthermore, human and financial resources are invested to maintain the chain of custody in order to maintain the integrity of investigation procedure [26]. In contrast to the current chain of custody approach, appropriate digital forensic investigation in large-scale IoT systems demands a more robust integrity preservation technique and faster methods.

As a result, in this research, we propose a blockchain-based digital forensic framework for the IoT environment to address the heterogeneity and spread of the IoT

environment, as well as the centralization of existing forensic investigations. At the moment, blockchain technology is the safest and most secure method of preserving data integrity. In addition, we describe a revised block structure and methodology for the proposed framework for study. We will investigate the execution time and time complexity of the proposed digital forensic investigation framework and apply it to actual digital investigations in the near future.

Our subsequent goal is to check more Internet of Things related devices, conduct various case studies, and impose the proposed model to real-world-based digital forensic inquiries. Thus, these goals will be assessed utilizing a variety of IoT digital forensic models based on environment. Furthermore, we want to execute various simulations in a variability of IoT-dependent scenarios, including smart cities, smart homes, smart industries, and smart autos.

## References

1. Kumar G, Saha R, Lal C, Conti M (2021) Internet-of-Forensic (IoF): a blockchain based digital forensics framework for IoT applications. *Futur Gener Comput Syst* 120:13–25. ISSN 0167–739X. <https://doi.org/10.1016/j.future.2021.02.016>
2. Sharma PK, Ryu JH, Park KY (2018) Li-Fi based on security cloud framework for future IT environment. *Hum Cent Comput Inf Sci* 8:23. <https://doi.org/10.1186/s13673-018-0146-5>
3. Hassan MU, Rehmani MH, Chen J (2019) Privacy preservation in blockchain based IoT systems: Integration issues, prospects, challenges, and future research directions. *Futur Gener Comput Syst* 97 512–529
4. Kebande VR, Ray I (2016) A generic digital forensic investigation framework for Internet of Things (IoT). In: 2016 IEEE 4th International conference on future internet of things and cloud (FiCloud)
5. Nakamoto S (2008) Bitcoin: a peer to peer electronic cash system. <http://www.bitcoin.org>
6. Harbawi M, Varol A (2017) An improved digital evidence acquisition model for the Internet of Things forensic I: a theoretical framework. In: 2017 5th International symposium on digital forensic and security (ISDFS)
7. Perumal S, Norwawi NM, Raman V (2015) Internet of Things (IoT) digital forensic investigation model: top down forensic approach methodology. In: 2015 Fifth international conference on digital information processing and communications (ICDIPC)
8. Zhang Y, Wu S, Jin B, Du J (2017) A blockchain based process provenance for cloud forensics. In: 3rd IEEE International conference on computer and communications (ICCC)
9. MacDermott A, Baker T, Shi Q (2018) IoT forensics: challenges for the IoA era. In: 9th IFIP International conference on new technologies, mobility and security (NTMS)
10. Cebe M, Erdin E, Akkaya K, Aksu H, Uluagac S (2018) Block4Forensic: an integrated lightweight blockchain framework for forensics applications of connected vehicles. Cornell University. arXiv preprint <arXiv:1802.00561>
11. Oriwoh E, Jazani D, Epiphaniou G, Sant P (2013) Internet of things forensics: challenges and approaches. In: 2013 9th International conference on collaborative computing: networking, applications and worksharing
12. Ryu JH, Sharma P, Jo JH, Park JH (2019) A blockchain-based decentralized efficient investigation framework for IoT digital forensics. *J Supercomput* 75. <https://doi.org/10.1007/s11227-019-02779-9>
13. Kouwen A, Scanlon M, Choo KKR, Le-Khac N (2018) Digital forensic investigation of two way radio communication equipment and services. *Dig Investig* 26:77–86

14. Tripathi E, Kumar U, Tripathi SP (2023) Comparative analysis of techniques used to detect copy-move tampering for real-world electronic images. *Int J Integr Eng* 15(4):201–225
15. Sharma PK, Singh S, Jeong YS, Park JH (2017) Distblocknet: a distributed blockchains based secure SDN architecture for iot networks. *IEEE Commun Mag* 55(9):78–85
16. Wang M, Wu Q, Qin B, Wang Q, Liu J, Guan Z (2018) Lightweight and manageable digital evidence preservation system on bitcoin. *J Comput Sci Technol* 33(3):568–586
17. Rivera J, van der Meulen R (2013) Gartner says the internet of things installed base will grow to 26 billion units by 2020. Stamford, CT
18. Henry R, Herzberg A, Kate A (2018) Blockchain access privacy: challenges and directions. *IEEE Secur Priv* 16(4):38–45
19. Tripathi E, Kumar U, Tripathi SP (2023) Image splicing detection system using intensity-level multi-fractal dimension feature engineering and twin support vector machine based classifier. *Multimedia Tools Appl* 82:39745–39763. <https://doi.org/10.1007/s11042-022-13519-2>
20. Ethereum private network platform (online). <https://www.ethereum.org/>. Accessed on 04 Sept 2018
21. Geth (online). <https://geth.ethereum.org/downloads/>. Accessed on 04 Sept 2018
22. Mist browser (online). <https://github.com/ethereum/mist>. Accessed on 04 Sept 2018
23. Chithaluru P, Turjman FA, Dugyala R, Stephan T, Kumar M, Dhatterwal JS (2024) An enhanced consortium blockchain diversity mining technique for IoT metadata aggregation. *Futur Gener Comput Syst* 152:239–253. ISSN 0167-739X. <https://doi.org/10.1016/j.future.2023.10.020>
24. Salama R, Turjman FA (2024) A description of how AI and blockchain technology are used in business. In: AIoT and smart sensing technologies for smart devices. p 15. <https://doi.org/10.4018/979-8-3693-0786-1.ch001>
25. Tripathi AK, Mishra S, Vasudevan SK (2024) Smart diabetic prediction: an intelligent iot-based diabetic monitoring system with stacked spatio temporal features-based multiscale dilated deep temporal convolutional network. *Sens Imaging* 25:2. <https://doi.org/10.1007/s11220-023-00446-1>
26. Ryu JH, Kim NY, Kwon BW, Suk SK, Park JH (2018) Analysis of a third party application for mobile forensic investigation. *J Inf Process Syst (JIPS)* 14(3):680–693

# Addiction-Based Tenuous Community Detection in Online Social Media



Zain Ul Abideen, Fakhar Shahzad, Bharati Rathore, and Tiansheng Yang

**Abstract** In the last decades, more social media users have been connected to the corrosion and decrease of face-to-face communication. Increasing the number of social network interactions, such as cyber-based relationship addiction and constant dependence on online social media, has been noted to become tenuous in real life. In this paper, we challenge that extraction of online behavior and real-life behavior gives us a tenuous relationship in real life. We conducted a problem-based scenario to collect data to find the range of people with social media addiction and vulnerable relationships. We use the degree and distance centrality closeness algorithm (DADCC) to find a community using social media, not live without online social media, and also find a community that uses social media but is not highly abdicated. However, both types of communities create a tenuous relationship with each other. We explore the importance of real life, but people's relationship with real life becomes tenuous, for example, by constantly checking Facebook updates or tracking people for hours. There is no direct conclusion to define whether a person is a social media addict; however, many doctors have found similar symptoms such as depression, anxiety, and psychological disorders, which are evidence of social media addiction. The main purpose of our work is to indicate that people's involvement in real life is much more beneficial as compared to online-based work.

---

Z. U. Abideen

Automotive Engineering Research Institute, Jiangsu University, Zhenjiang, Jiangsu, China  
e-mail: [1000006198@ujs.edu.cn](mailto:1000006198@ujs.edu.cn)

F. Shahzad

Research Institute of Business Analytics and Supply Chain Management, College of Management, Shenzhen University, Shenzhen, China  
e-mail: [fshahzad51@szu.edu.cn](mailto:fshahzad51@szu.edu.cn)

B. Rathore · T. Yang (✉)

University of South Wales, Pontypridd, UK  
e-mail: [tiansheng.yang1@southwales.ac.uk](mailto:tiansheng.yang1@southwales.ac.uk)

B. Rathore

e-mail: [bharati.rathore@southwales.ac.uk](mailto:bharati.rathore@southwales.ac.uk)

**Keywords** Addiction · Social media · Tenuous community · Degree and distance centrality crossness · Digital detox

## 1 Introduction

Since early, social scientists used the concept of social networks. In 1954, John Arundel Barnes initially used the term systematically to express the ties and relationships among nodes, social groups, and social categories. The social network seems to be the natural way to organize individuals, resources, and interactions in an effective and robust structure [1]. Therefore, one way to use this data is to compose networks of relationships between individuals or objects, such as mapping friendships among Facebook users, tracking purchasing patterns of Amazon customers, or analyzing connections between users and genetic products. The field of community detection aims to find objects or individuals connected with each other inside the networks. All of these groups are called communities. The main motivation of the field of community detection is to understand the opinions of different groups of people or influence among people. The field is also useful for detecting the e-commerce web to build the way of purchasing different people in different states. Social network analysis (SNA) is a process of investigating social structure using graph theory and networks. We can characterize the whole network structure in terms of nodes. These nodes consist of the individual actor, things, or peoples within the networks and their ties, relationships, edges, or interactions between the connections. However, the examples of the social structure, usually visualized through SNA, included social media networks. All of these tasks are spread through information circulation, friendship, and acquaintance networks, social and business networks collaboration graphs, sexual relationships, and disease transmission. Nowadays, a lot of work has been conducted in community detection, but finding socially tenuous groups is less explored in the community determination area [2]. The social network is very wide and popular today, and the accessibility of social media is very easy and high. There is no comparative psychological study of the uncertainty; the experiment shows access to social media becomes uncertain in human life [3]. We access social media through Facebook, LiveJournal, LinkedIn. The research on finding several social groups for community detection and activity coordination is increasing attention daily [4, 5]. Therefore, socially connected individuals in dense groups from online social networks are identified by extracting these types of groups. However, socially tenuous or weak groups can have a negative impact on real-life and psychosocial outcomes, including depression, anxiety, severe isolation, and, tragically, even suicide. An example of this tenuousness in real life can be seen in the case of a person with 1,250 Facebook friends, 588 Twitter followers, and 743 WhatsApp contacts, yet in the moment of crisis, such as being in the intensive care unit (ICU), the only people present were his wife, children, and parents. For whom never had time to spend. In this paper, we use social media abdication to impact on real life. In real life, we have a tenuous relationship with people. In order to gather datasets, we conducted a survey of 6 questions to check whether people

are social media infected, less infected, or highly infected. Social media addiction spread some weaknesses:

- Lacks emotional interaction.
- Decrease face-to-face (F2F) communication skills.
- Gives people a license to be hurtful.
- Carries inauthentic appearance of feelings.
- Facilitates laziness.
- Reduces understanding and thoughtfulness.
- Decreases family closeness.
- Create a twisted self-image.

In this paper, we discussed social media network dependency therapy. This kind of therapy determined the tenuous relationship between real life and online social life. We conduct a survey in China and Pakistan universities. The basic agenda of this survey is to determine the social media addiction community. This type of community is where people sit together, like meet family, friends, colleagues, and other ceremonies, and most of the time, they use mobile phones, tablets, or PCs. In this situation, the peoples are tenuous with respect to each other in real life. We conducted the survey based on the following questions: We examined the percentage of people who are addicted to social media and become a tenuous group in the community in real life.

- Do you think about social media or plan to use social media a lot while offline?
- Do you feel a craving to use social media more and more as days pass?
- Do you use social media to escape from personal problems?
- Have you tried to quit social media more than a few times, but every time you failed miserably?
- Can you say that you frequently make attempts to cut back on your social media usage and accomplish little?
- Do you feel restless or upset if forced to go without social media use?
- Do you find yourself checking social media many times a day and it is getting in the way of your work, studies, or relationships?

On behalf of these questions, if your answer is yes to a few of these questions, you are impartially standard, usually social media users. We have a way to use less social media, digital detox is a way. The digital detox is a way of pushing us to spend less time on social media so as we can indulge in real life for a change. Comment below your most used way to detox in real life and turn off push notifications, convert it black and white, put away the phone during meal time designate tech-free hours at a dinner table, or make bed room a no-tech zone. Limit yourself to one screen usage (no multitasking), spring cleaning social media accounts, choose right apps for download, etc.

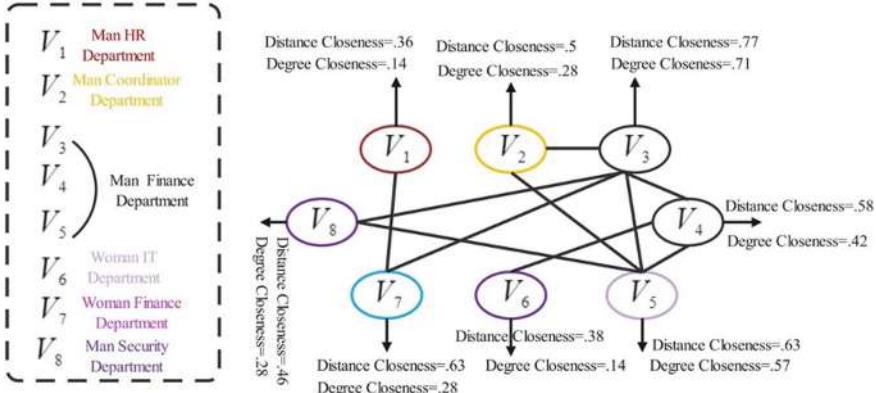
We can protect ourselves and our bodies by adopting these simple ways in real life. There are following contributions are:

- Nowadays, social media addiction is commonly managed at a late stage. To address this problem, we designed and proposed a method to detect the community has a tenuous relationship.
- We develop a method that efficiently and reliably shares intermediate outcomes across centrality computations.
- We present a method for computing degree closeness and distance closeness centrality that minimizes estimate completion.
- Through distance closeness centrality, we design a pruning-based technique, which uses machine learning approach.
- We experimentally reveal the benefits of the above features using real-world networks.

The paper is organized as follows. Section 2 formulates the concept of a social media dependency group, examining how individuals' reliance on online platforms can influence their social behavior, mental health, and real-world relationships. Section 3 presents a review of related work, highlighting previous research and key findings that contribute to the understanding of the topic at hand. Section 4 has a roadmap of the problem statement and algorithms for seed selection and pruning based on distance centrality closeness on graphs. Section 5 illustrates the empirical study, comparison, and evaluation, and Sect. 6 has the conclusion and future work.

## 2 Social Media Dependency Group Formation

How we imagine and subjectively experience the future can inform how we make decisions in the present [6]? This has been an assumption in the field of psycholinguistic research, concerned with how adults process incoming language online [7]. According to group therapy for constituent abuse treatment, for group formation cognitive behavioral or psychoeducational is an important task [8]. The principle of forming the hopefully socially tenuous group to whom unknown patients are allocated on this purpose must ultimately be individual selection with phenotyping similar disorder signs and behaviors. Our purpose for engaging members is to avoid using mobile phones or tablets; the members communicate with each other without any hesitation. Moreover, we find subgroups of members in the tenuous group which act against other peoples during the therapy session. According to the scenario, a clinical psychologist would like to visit a company's annual dinner. The psychologist selects a table, where 8 people sit together, 3 of females and 5 of men. Out of 8 people, 3 belong to the same department, and the remaining belong to different departments. The clinical psychologist observes that some of the members do nothing; they only look around; some of the members continuously use their mobile phones, some of them take photos of the dinner party or take their own selfies, and again, they are busy using the mobile phone. After observation, the psychologist added the 8 people. The phycologist said I asked some questions of all of you. All of you give answers according to their situation. The phycologist makes a graph and assigns the degree of



**Fig. 1** Motivative example with value of distance and degree closeness

each node according to the yes or no of the individual, giving answers to the question. In the end, the psychologist finds the degree of closeness of each member on behalf of maximum attachment with social media. The psychologist makes subgroups on behalf of their answers. The phycologist selects the members who need a digital detox strategy. The remaining members and social media-infected members are tenuous with real life. Please do not alter the formatting and style layouts that have been set up in this template document. According to questions-based answers, these nodes belong to different gender and belong to different department. The same department gives one edge to become familiar with each other as compared to others. Subgroup consists of nodes  $F1 = \{V3, v4, V5\}$  is not tenuous nodes with respect to social media addiction. So  $F2 = \{V1, V2, V6, V8\}$  tenuous nodes, as these individuals engage less with social media and therefore do not require digital detox-based therapy. Therefore, the physiologist chooses  $F = \{V3, V4, V5, V7\}$  for therapy.  $V7$  also chooses less but has a direct relationship with these nodes because  $V7$  also belongs to the HR department, so  $V7$  is familiar with  $V3$ ,  $V4$ , and  $V5$ . Remaining nodes are  $V1$ ,  $V2$ ,  $V6$ , and  $V8$ . These are tenuous in real life with respect to social media addiction. But social media-addicted people, and these people have no communication with each other. Make a tenuous relationship with each other (Fig. 1).

### 3 Related Work

Concerning many social applications, extraction of dense subgraphs or communities is a very important to community detection. Finding the community, there exist serval methods such as dense subgraphs extraction; several social cohesive measures have been proposed; examples are density [9], diameter [10], truss, clique, and its dissimilarities [11]. The extraction of densely connected subgraphs from social networks has been actively studied and holds significant importance in community detection

methods [12]. The peoples are densely connected with each other through social media. We aim to extract groups of people like subgraphs. Research organizing and scalability have a lot of importance, on the behalf of organizing social groups based on the tightness among standing friends and further crucial factors have also been studied [13]. Therefore, GSGQ [14] and MRGQ [15] both extract socially dense groups with 3D constraints; we design the algorithm to examine the distance closeness and degree maximum of the nodes. These algorithms are faster to implement and easy to explain how to implement. Recently, research has focused on the sampling of graph networks and scarification, spanners, and simplifications [16]. In this paper, we design an algorithm to find out the degree closeness and distance closeness centrality. We define the predefined value of the parameter. We change the value of the parameter, crucial impact change with the increase or decrease of the value. As before, TERA, TERA-ADV [17], and Noah Ark Principle [18] are used to find subgroups that are unfamiliar constraints. This algorithm in prior work cannot be applied, because finding socially tenuous groups' application have less research interaction. We design distance closeness and degree centrality algorithm to find out unfamiliar subgroups in the better way. As prior work, the  $K$ -densest subgraph problem work Sotirov et al. [19], to find a subgraph of  $k$  vertices, uses the chordal graph with the least induced edges. Moreover, Fuster et al. [20], proposed a promising approach to identify groups of users with the highest similarities, despite the fact that most people in social networks do not know each other. Bai et al. [21] give the novel concept of minimum triangle disconnection to find out the tenuous groups. But he uses the very simple graph to draw their results. We used the attributed network to find out the tenuous groups in real life vs. online internet life. We further improve the distance selection and more improve the results. We conduct a survey of some question related to social media addiction, make life so tenuous and alone. With the world currently in an age of increased mobility, it becomes crucial for us to understand how people move from place to place so we can make patterns about where a person visits often between one region or another, and what kind of words are relevant when such places appear. We find that the number of PoIs is constrained by properties shared among individuals [22]. The frequent broadcasting of media reports on moral crises like famine can actually dampen as much as spur moral concern. They failed because people assumed, often incorrectly, that they had seen all the news and formed their opinions based on limited information. Some of us may care less, or have never seen or heard about the issue at all [23]. We check all the percentages of social media-addicted people; we also use datasets from Sanford websites and implement our algorithm to find out the subgroup's tenuity. The peoples are social media addicted, but they have no idea about this. Through this paper, we also told the positive way of life, balanced life. Proper use of online and offline. We use digital detox as a way to balance our lives. Some steps to achieve this include: seeking help if needed, relearning how to entertain yourself without the glow of a screen, reconnecting with old friends, using pen and paper, hiding your smartphone, switching to a simple watch instead of a smartphone, spending time in places where cellphone use is discouraged, stopping unnecessary email checking, and scheduling time for a break on your calendar.

## 4 Problem Statement

Given the undirected graph of social network  $G = (U, V, E)$ . Let  $U$  and  $V$  denote the number of vertices and neighbors with respect to each other. The DADCC problem is formulated as follows:

**Given:** The social network of  $G = (U, V, E)$ , the graph's size constraint is  $n$ , and the tenuity parameter is  $\varepsilon$ .

**Find:** We find a subgraph  $G' \subseteq G$ , where vertices  $U, V$  have the maximum value of degree centrality and distance closeness centrality of the vertices  $U, V > \varepsilon$  have no-pair constraint such that we minimized the pair connectivity. The group is at least the average value of distance between any two nodes in the original graph.

**Definition 1** Given an undirected and unweighted graph, the  $k$ -hop distance, the size of the constraint  $n$ , and the tenuity parameter  $\varepsilon$ .

**Definition 2** In the given graph, we have a user-defined threshold parameter  $\varepsilon$ . We compare the value of the threshold with our distance.

**Definition 3** We count the sequence of the consecutive pair from  $u$  to  $v$   $(u, v_1, v_2, v_3, \dots, v_k, v)$ , where  $v_1, v_2, v_3, \dots, v_k \in V / d(u, v_1), d(u, v_2)$  compare with the threshold. We design pruning techniques using this threshold value.

**Definition 4** We calculate the vertex  $u$  degree closeness with degree vertex  $v$ . So  $C_{\deg}(V) = \frac{d_v}{N-1}$  where  $N$  is the set of nodes in the given graph network, while  $d_v$  is the degree of node  $v$ .

**Definition 5** We use distance centrality closeness of the shortest path from vertex  $u$  to vertex  $v$ .

$$C_{(u)} = \frac{n-1}{\sum_{v=1}^{n-1} d(v, u)} \quad (1)$$

The distance closeness centrality of node  $u$  is the reciprocal of the  $av$ . Distance to  $v$ , all reachable  $n - 1$  nodes.

**Definition 6** We find a subgraph, in which subgraph  $H \subseteq G$ . Therefore,  $|H| \geq n$ ,  $n$  is a size constraint. In subgraph  $H$ , edges of  $u$  and  $v$  have a no-pair constraint, so we minimized the frequency factor (Table 1).

### 4.1 Personalized Recommendation

We have an undirected graph, where  $V = \{1, 2, 3, N\}$  is the vertex set,  $E$  is the edge set, and  $E \subseteq V * V$  of the graph  $G$ . In the graph  $G$ , we have  $(u, v)$ , we use edge from  $u$  to  $v$ . We design an algorithm to allocate a path  $Pv0$  where  $vp = (v0, v1, v2 \dots Vp)$

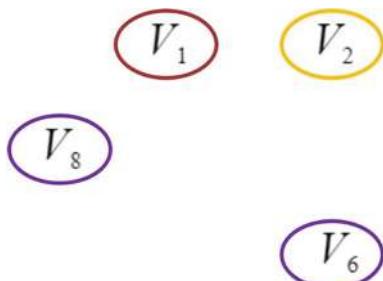
**Table 1** Frequently used symbols

Symbol	Interpretation
$G$	Undirected graph
$V$	Set of vertices
$H$	Subgraph of $G$
$N$	Size of constraint
$E$	Set of edges
$U, V$	Pair set of vertices
$\varepsilon$	Predefined threshold
$d(u, v)$	Distance from vertices $u$ to $v$
$C_{\text{deg}}(V)$	Degree closeness of vertex
$C_{(u)}$	The av. distance closeness of vertices
$F$	Allocation for groups

represents a simple path from vertex  $v_0$  to  $v_p$ . The number of edges between a vertex and the next vertex in an undirected graph determines the length of the path in the graph. From vertex  $u$  to vertex  $v$ , we denote distanced as  $d(u, v)$ . The minimal length depends on all paths from  $u$  to  $v$ . For example, according to Fig. 2, we know the degree centrality  $C_{\text{deg}}(v_i)$  of each node. We normalized degree centrality  $\frac{C_{\text{deg}}(V)}{n-1}$ ; on the behalf of normalized degree centrality value, we choose the value of predefined threshold  $\varepsilon$ , whether our parameter  $\varepsilon$  is greater than or less than? However, initially, we design Algorithm 1 to explore every node and label each node with its degree centrality value  $C_{\text{deg}}(V)$  as well as its normalized degree centrality  $\frac{C_{\text{deg}}(V)}{n-1}$ , where  $n$  is the total number of nodes. Using Algorithm 1, we select a seed to save time and predict the behavior of nodes during graph iterations, *gt* mention graph in it a rational phase, in this way we calculate the degree centrality  $C_{\text{deg}}(V)$  for each node in the graph, and then normalize the value of each node using  $\frac{C_{\text{deg}}(V)}{n-1}$  where  $n$  is the total number of nodes.

This process determines the value of  $\varepsilon$  for maximum degree nodes. However, we use the distance closeness algorithm for pruning and to remove direct relationships between nodes, thereby fostering more familiar relationships. In Fig. 2 we mention

**Fig. 2** Tenously isolated group



the value of distance centrality closeness and degree centrality of each node. All the values are also given in tabular form. In the first algorithm, both  $u, v$  are considered active. The meaning of action is all the vertices begin to participate in computation. Moreover, it is mentioned that hyphenation should be avoided at the end of a line. We embolden the symbols of vectors and matrices. Scalar variable names are to be italicized. All weights and measures shall be given in SI units. Nonstandard abbreviations or symbols should be defined the first time they appear, and a glossary is required for each other. We use a pruning-based algorithm in which the node has a distance  $> \varepsilon$ . Furthermore, if we change the value of  $\varepsilon$ , the crucial impact is to change with respect to change the value of  $\varepsilon$ . The best approach to selecting the appropriate value of  $\varepsilon$  is to take the average of all distance closeness centrality values. The average value helps us to set the  $\varepsilon$  as exact to prove error-free. For example, in Table 2, we have values of distance centrality; we take the average of these nodes, set the  $\varepsilon$ , and remove all the nodes that have greater values as compared to  $\varepsilon$ .

$$\text{Av. } C_{(u)} = \frac{n - 1}{\sum_{n=1}^{v=1} \frac{d(u,v)}{n}} \quad (2)$$

$$0.36 + 0.5 + 0.77 + 0.58 + 0.63 + 0.38 + 0.63 + 0.46/8 = 0.5 \\ 0.36 + 0.5 + 0.77 + 0.58 + 0.63 + 0.38 + 0.63 + 0.46/8 = 0.$$

#### Algorithm 1: Seed Selection

**Input:** an undirected graph  $G = (u, v, e)$

**Output:** In the graph, labialized all the vertices with normalized degree centrality values

1.  $u, v.Active = True \forall u, v \in g$
2.  $\varepsilon \leftarrow 0$
3.  $t \leftarrow 14gt \leftarrow g$
4. *if*  $C_{\deg}(V) = \text{degree of node}$
5. *then*  $C = \frac{C_{\deg}(V)}{n-1}$
6.  $C \leftarrow \text{Max\_Selection}(u, v)$
- $\forall u, v \in gt$
7.  $C > \varepsilon$
8. *return*  $\varepsilon$

**Table 2** Mathematical analysis of DADCC algorithm

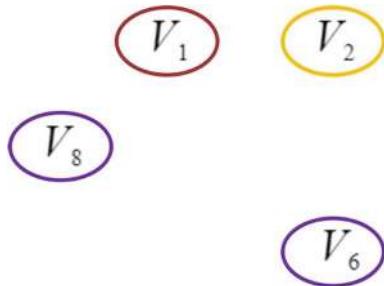
Node	Distance centrality closeness	Degree centrality	Normalized degree centrality
V1	$7/0 + 3 + 2 + 3 + 3 + 4 + 1 + 3 = 7/19 = 0.36$	V1 = 1	$1/7 = 0.14$
V2	$7/3 + 0 + 1 + 2 + 1 + 3 + 2 + 2 = 7/14 = 0.5$	V2 = 2	$2/7 = 0.28$
V3	$7/2 + 1 + 0 + 1 + 1 + 2 + 1 + 1 = 7/9 = 0.77$	V3 = 5	$5/7 = 0.71$
V4	$7/3 + 2 + 1 + 0 + 1 + 1 + 2 + 2 = 7/12 = 0.58$	V4 = 3	$3/7 = 0.42$
V5	$7/3 + 1 + 1 + 1 + 0 + 2 + 2 + 1 = 7/11 = 0.63$	V5 = 4	$4/7 = 0.57$
V6	$7/4 + 3 + 2 + 1 + 2 + 0 + 3 + 3 = 7/18 = 0.38$	V6 = 1	$1/7 = 0.14$
V7	$7/1 + 2 + 1 + 2 + 2 + 3 + 0 + 2 = 7/11 = 0.63$	V7 = 2	$2/7 = 0.28$
V8	$7/3 + 2 + 1 + 2 + 2 + 3 + 2 + 0 = 7/15 = 0.46$	V8 = 2	$2/7 = 0.28$

## Algorithm 2: Pruning based on distance centrality closeness

**Input:**  $G = (u, v, E), n, \varepsilon$ **Output:** Pruning the node has greater value  $\varepsilon$ 

- 1:  $Gt \leftarrow G, i \leftarrow 1, U \leftarrow \emptyset$
- 2: *While*  $|n| > \varepsilon$  *do*
- 3: We identify the value of  $\varepsilon$ , maximum degree nodes have a greater value of  $n$  than  $\varepsilon$ . So, we remove all the node have a higher value than  $\varepsilon$
- 4:  $Gt + 1 \leftarrow Gt - \{vi\}$
- 5: Therefore,  $Gt + 1$  satisfies no-pair constraint then
- 6:  $U \leftarrow U \cup \{Gt + 1\}$
- 7: *end if*
- 8:  $i \leftarrow i + 1$
- 9: *end while*
- 10:  $Gt \leftarrow G$
- 11:  $Gt \leftarrow \text{Attributed selection measure}$
- 12: *end if*
- 13: output  $Gt$

The  $\varepsilon = 0.5$  removing all the nodes that have a greater value as compared to  $\varepsilon$ . Figure 3 shows the nodes that are tenuous. This node does not have familiarity with each other's. Other nodes are highly infected with social media addiction as compared

**Fig. 3** Tenous subgroup

to these nodes. We design pruning-based algorithm 2, in which we define the value of  $\varepsilon$ , take the average of distance centrality closeness of nodes, and set the higher-degree nodes values and lower degree nodes values. We observe the higher-degree nodes which have greater values than  $\varepsilon$ . In algorithm n and  $\varepsilon$  are parameters; on the behalf of these parameters, we determine the tenuous nodes from the community.  $Gt$  is a graph in which iteration computation implements.  $I - 1$  is the number of iterations, which goes to  $k$ th iteration. At  $Gt + 1$ , the pruning phase begins, and the removal of nodes starts.  $Gt - \{vi\}$  represents the remaining graph after the node  $vi$  is removed. We extract the subgraph  $Gt$ , but one thing is necessary, we calculate the attributed selection measurement. Attribute selection is the process, in which we are removing the redundant attributes. These redundant attributes are irrelevant to the mining task. There are some different steps which are given below as:

**Information Gain:** Information gain is the total amount of information, which is collected by knowing the values of the attribute. The entropy of a distribution is the difference between the entropy before the split and the entropy after the split. The largest value gives the smallest entropy.

$$\text{Info } (C(u)) = -n - 1 / \sum n - 1v = 1d(v, u) \log 2d(v, u) \quad (3)$$

**Gain Ratio:** Information gain gives many outcomes on behalf of their results. Therefore, information gained is used for partitioning the maxima of the attributes.

$$\text{Gain ratio}(n) = \text{Gain}(n) / \text{Split info}(C(u)) \quad (4)$$

**Gini Index:** When we calculate the gain ratio, for removal of any impurity we use Gini index.

$$\text{Gini } (C(u)) = 1 - \text{sigma } m.i = 1pt2 \quad (5)$$

$P$  is use for probability of the  $C(u)$ .

**Error Complexity Pruning:** The main concern of error complexity pruning is the calculation error cost of each node. It finds the complexity of each node. The following equation is useful for calculating this kind of error:

$$R(t) = r(t) \quad (6)$$

$r(t)$  is the error rate of a node. Therefore  $r(t) = \text{no of examples of misclassified in nodes/no. of all examples in the node.}$

$p(t)$  is the probability of the node.

$p(t) = \text{no. of the examples in node/no. of the total examples.}$

**Minimum Error Pruning:** The following equation used for the prediction about expected error rate of pruning at node t.

$$E(t) = nt - nt, c + k - 1/nt + k \quad (7)$$

where,  $k$  is mention number of classes,  $nt$  mention no. of examples in node  $t$ ,  $nt, c = \text{no of example assign to } C(u)$ .

At each node in the dense subgraph, the calculated error rate in the subgraph pruned. Calculate the expected error rate for the subgraph. The results of all mathematical calculations show in the experimental part.

## 4.2 Distance and Degree-Pruning

From the inspection and evaluation of DADCC, we find that it is unnecessary to scan over all vertices in  $g$  due to many will not meet no-pair constraints. Some of the vertices may be unnecessary and can be removed from the graph. In DADCC, we design pruning-based techniques, pruning based on preprocessing and post-processing techniques. In our proposed solution, we explore the nodes with degree closeness values during the preprocessing phase. In the second algorithm, we calculate the distance closeness centrality, and reduce the size of the vertex sets post-pruning. In both processes, we gain information; these procedures are very useful during the runtime. Here we express some lemmas, these lemmas analytically satisfy the algorithms, in which we associate the parameter  $\varepsilon$ .

Given a graph  $G = (u, v, E)$ , let  $u, v \in x$  vertex, Let  $N(x)$  we mention the  $1 - \text{hop}$  distance from vertex to its neighbors, Let  $N[x]$ , mention as a close neighbor, the closed neighbor have a distance closeness is greater than  $\varepsilon$ , i.e.,  $N(x) = N(x)U\{x\} > \varepsilon$ , seed selection in algorithm 1. We have tenuous vertex  $s, s \in u, v$  to form a subgraph, less than value  $\varepsilon$ . For an efficient and reliable solution, obtained a feasible solution  $F$ . The  $F$  define and satisfying the constraint less than  $\varepsilon$ , have no-pair link with each other.

**Lemma 1** Given the required achievable solution  $F$  and the tenuous vertex  $s$  have  $|F \cap N[s]| < \varepsilon$ .

**Proof** The vertex  $s$  is tenuous on the behalf of  $\varepsilon$ , and  $N(s)$  form a subgraph where  $(u, v) \in E$  for any  $u, v \in N[s]$ , if feasible solution  $F$  has an edge, then as achievable (feasible) solution  $F$  is not an existing solution. Therefore, for tenuous vertex  $s$ , we select the vertex in  $N[s]$  trim all other vertices have distance closeness value greater than  $\varepsilon$  according to algorithm 2, to ensure not have a direct relationship. Which trim satisfies the no-pair constraint and generates the minimal objective value.

**Theorem** Therefore, for any feasible solution  $F$  gained from  $G$ , there exists a solution  $F'$  in degree closeness-based pruning no worse than  $F$ . Moreover, if  $|\varepsilon| < n$ , there is no feasible solution to the problem. This theorem is associated the pruning.

**Theorem** Therefore, for any feasible solution  $F$  gained from graph network  $G$ , there exists a solution  $F'$  in degree closeness-based pruning no worse than  $F$ . Moreover, if  $|\varepsilon| < n$ , there is no feasible solution to the problem. This theorem is associated with pruning.

**Proof** Let  $F$  be a feasible solution obtained from  $G$  and assume  $N(s)$  and  $x \in F$ , where  $s$  is  $x$ 's corresponding  $s$  tenuous vertex. Based on Lemma 1, is a better solution than  $F$ . Therefore, there is always a solution  $F$  in solution no worse than  $F$ .

Therefore, if  $|\varepsilon| < n$ , so we suppose that the set of vertices in  $\varepsilon$  has the values based on  $av.$  shortest distance. However,  $\varepsilon = sUw$ , where  $w$  is the set of vertices that have the value less than  $\varepsilon$ , but have the familiar ratio, neither exists nor connected. The distance of  $d(u, v) = G - w$  holds. If  $s$  has an edge with vertices, this violates the pruning condition. So therefore, there is no solution. In the following, the graph is linear if for any two vertices  $u, v$  in the graph,  $u < \varepsilon$  or  $v < \varepsilon$ , or both. The following lemma in the work first expresses the one-to-one correspondence between a threshold graph.

**Lemma 2** A graph  $G$  is a threshold graph if and only if the vertices  $u$  and  $v$  are adjacent in  $G$  is linear.

**Lemma 3** According to this lemma, the vertices in the graph  $G$  can be labeled such that  $d(u_1 \dots u_n) + d(v_1, v_n)\varepsilon|G|$ . Therefore, for the optimality of algorithm, we will first explore the value of  $\varepsilon$  through average shortest path distance in a threshold graph.

**Lemma 4** Given a threshold graph  $G = (u, v, e)$ , for any three vertices  $u, v, x \in d$ , if  $u \sim v$ , then  $dG(v, x) \leq dG(u, x)$  hold.

**Proof** There are two cases.

1. if  $v$  is at the shortest path from  $u$  to  $x$ , represented as  $PG(u, x)$ , then  $dG(v, x) < dG(x, u)$  holds. This case holds seed selection algorithm 1.
2. If the  $v$  is not at the  $PG(u, x)$  then there exists a vertex  $v \in N(x)$ , and  $v \in PG(u, x)$ . Since  $v \in N(x)$  implies that  $v \in N(v)$   $dG(v, x) \leq dG(u, x)$  holds for the reason that the path  $PG(v, x) = \{v\} \cup \{u\} \cup dG(v, x)$  must have length no larger than that of  $PG(u, x) = \{u\} \cup \{v\} \cup dG(v, x)$ . This case holds trimming algorithm 2.

### Time Complexity Analysis for Pruning-Based Algorithm:

Let be denoted maximum degree as  $D$  of the vertex in the given graph. If the vertex  $x$  has degree closeness, it requires  $O(D2)$  time.  $O(D2d(u, v))$  time, and distance closeness require  $O(d(u, v)2)$  time. Therefore, the overall time complexity of our DADCC is  $O(D2|d(u, v)| + |d(u, v)|2)$ .

## 5 Experimental Evaluation

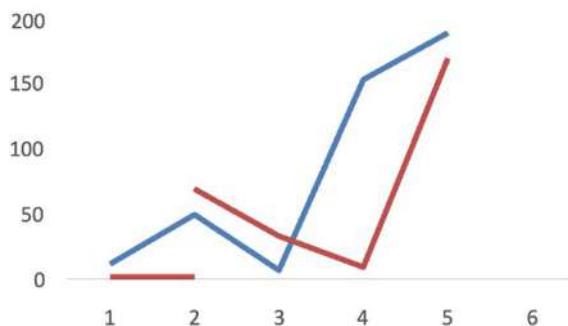
Our experimental part consists of two sections, one part we conduct a survey-based question section and second part we take datasets of different social networks and implement an algorithm. In study 1, the researchers sought to determine whether there is a group malleability reality link. To this end, at Xian Jiaotong University (XJTU) Xian City Shaanxi Province China, students of this university are the participant. The questions about social media addiction and real-life isolation ask by the students [24]. The nature of the group is malleable. There are 200 students, consisting of 60 females and 140 males. This group of a student is a community, approached on the social media-based question survey. The age of the participants ranged from 20 years to 25. Other factors include age, gender, education, frequency of social network site usage, and duration of each visit to the site [25]. As a part of the survey, participants students were first asked to complete the answers of the survey-based papers. They were presented with a short question given in the introduction part. The participants' reactions to the collective use of social media across various devices, including mobile phones, tablets, PCs, etc. After finishing the study, all subjects were extensively debriefed on involved in this study. As such, participants received lay language descriptions of the psychological constructs examined in this study and how those constructs were expected to relate. All the variables measured on the response of the participants from 1 (strongly agreed) to 6 strongly disagreed.

All of the data use as datasets. Find the value of  $\varepsilon$ . The students vary under the value of  $\varepsilon$ ; yes or no answers change the value of  $\varepsilon$ . But the young generation is most addicted to internet social media and mental disorders. According to Fig. 4, this is the survey of Xian Jiaotong University 200 students. In real life, peoples are social media addicted. In the pic, the y-axis represents participants, and the x-axis represents questions asked for agreeing or disagreeing. Males and females have little bit different to give answers yes. People need digital detox therapy. To identify whether a group's malleability is affiliated with real life. To this end, at Comsats University Islamabad Pakistan, students at this university is the participant. The questions about social media addiction and real-life isolation ask by the students. The nature of the group is malleable, 450 students in which 120 females, 330 males. This group of a student is a community, approached on the social media-based question survey. The age of the participants ranged from 23 years to 35.

As part of the survey, student participants were first asked to complete the survey questionnaire. They were presented with a short question give into introduction part.

**Fig. 4** XJTU students survey

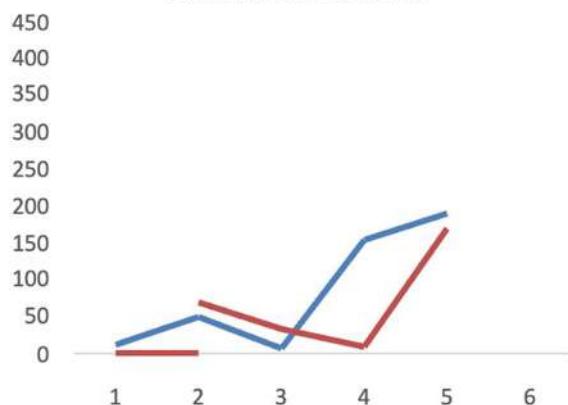
Xian Jiaotong University Survey between 200 students



The participant's reactions to this impending collective to the use of social media, Mobile, tablet, Pc etc. At the end of the study, participants were debriefed in full about what we wanted to find out. As a result, in the information session participants were presented with plain language explanations of the psychological variables under investigation and what is thought to be their relationship. All of the variables were responded to by a participant from 1 (strongly agree) to 6 strongly disagree. All of the data is used as data sets. Find the value of  $\varepsilon$ . The students vary under the value of  $\varepsilon$ , yes or no answer change the value of  $\varepsilon$ , But young generation most addicted by internet social media and mental disorder. Graphically result as shown above. According to study to Fig. 5, this survey conducts in Pakistan, most of the student's half of the students agree and half of the student disagree. Half of the students did not attempt some questions. As compared to China, Pakistani students are used to mobile, tablets, or pc, all things related to social media addiction.

**Fig. 5** Pakistani university-based survey

Comsats University Survey between 450 students



**Table 3** Evaluation and comparison

Algorithm name	Attributed	Results improvement (%)	Feasibility ratio (%)	Objective value ratio (%)
TERA	No	70–85	55–92	50–90
TERA ADV	No	80	60–90	60–90
UTNA	Yes	75–91	70–90	50–80
DADCC	Yes	85–95	75–95	70–95

### 5.1 Evaluation and Comparison

In this section, we will compare our proposed solution DADCC with prior work. Our solution is 10 to 20% improved as compared to prior solutions. We have maintained accuracy and attributed graph. Also maintain the feasibility and objective value on different datasets. Our method uses supervised learning. In which we select seed and trim all possible edges, which creates pair constraint. For better evaluation and comparison, we discussed as tabular form as Table 3.

We use real datasets in our algorithm. Before we apply to the 2 universities, one in China and 2nd in Pakistan. Now, we apply it to large datasets. In China, some social networks are banned, so we use around the world datasets for our experimental approach.

### 5.2 Data Collection and Evaluation

Here, we describe the datasets we gather to produce accurate information about isolation (Table 4).

As a case study, research into human decision-making has time and again neglected to consider the question of a search although it must play an important role in many domains including but not limited to foraging or hunting, mating or

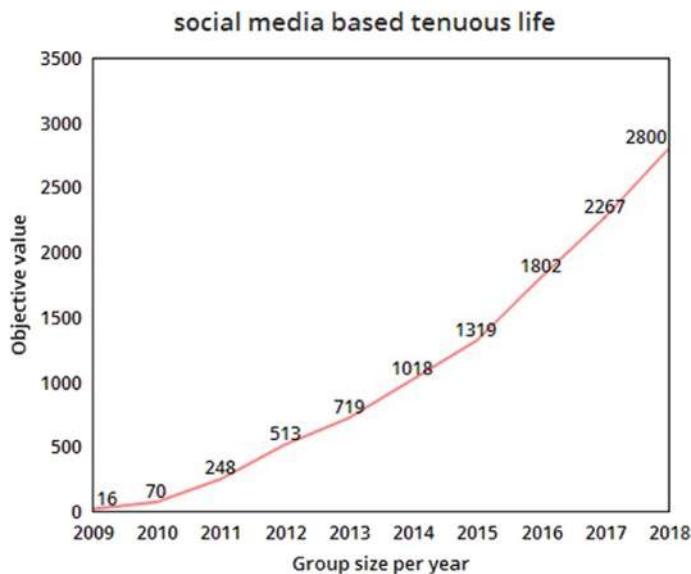
**Table 4** Datasets description

Datasets	Datasets illustration
FB-US	The friend of a user, user profile, user-created news-feeds/wall in which likes a comment or tagging users also included. Events join or decline. Groups joined w.r.t to events. Game notifications, requests created by game applications
IG-US	User profile, each user follower's user-created news-feeds in which likes a comment or tagging users also included and the users like or comment based contents
FB-L	User ID anonymized w.r.t to perform the action, also user ID anonymized that gets the action. Times. Amp of action creation
IG-L	ID of the anonymized user who takes an action and ID of the anonymized user who receives this action. Tags (the entire tag set assigned to the media), likes, comments

forming alliances, visual target acquisition as well finding information [26], for reasonable accuracy, we rely on the account name and their creation time. We recruit 3500 online social media worldwide users via Amazon Mechanical Turk (Murk) for the classification of tenuous community datasets obtained for the training and testing process. The participants included 1878 men and 1622 women. Therefore, the professions of all the participants are very diverse. Affiliated with various professions, such as universities, government offices, art centers, technology companies, businesses, and banks. All questionnaires are to be completed by each user first. More specifically, we are interested in perceptions of political leaders and celebrity actions on social media: the extent to which they believe that leaders/celebs behave with intentionality as determined by their online activity. Then a group of professional psychiatrists participating in the project asses, label to the user's potential cases. We find the community with attributed datasets are tenuous, and we use the datasets of crawls as Facebook denoted as FB-US, Instagram IG-US, all of the information gathered with the Facebook and Instagram APIs, these data mentioned in tables. During the experiments, we determine the impact of relationships and extract social interaction according to addictions. We compare the proposed algorithm method of DADCC with other baseline approaches. We observe that some previous algorithms do not scale up the large networks, such as Brute force with all possible combinations. We design small groups of datasets FB and IG and compare the results. According to analysis, on some large datasets of online social networks, in Fig. 3a and b, left most sideshow all users, 45% friends are CR users. On the other hand, the 9th bar Fig. 3a indicates FB-L about 60% of users. Figure 3a and b illustrates that FB-L, IG-L, CR, and IO all are the same as friends. CR and IO by nature are identical.

In Fig. 6, we use the datasets of the last ten years, we gain information from 2010 to 2018, and the number of peoples increases day by day using online social media. Peoples spend a lot of time on the internet or online activities, but not have time to meet in real life. Furthermore, FB comprises 64 K vertices and 818 K edges, and IG contains 46 K vertices and 679 K edges [26]. The 3rd dataset, DBLP, is a co-author infrastructure with 317 K nodes and 1.1 M edges 8.4. The fourth dataset is collated with the network with 1.7 M vertices and 30 M edges 9. In our experiments, our default parameter is  $k = 0.5$  and  $n = 20$ . The algorithm is implemented on Dell DL 4510 server with core i5 3.0 GHz and 1 TB ram. Therefore,  $k$ -hop graph,  $0.5 \leq k \leq 0.7$ . The input graph is fileted by DADCC algorithms. Compare the feasibility ratio with the different sizes of constraint, we get to obtain a feasible solution tested for a different interval of time. There is a table is given below, in which we use the 4 datasets, their vertices, and edges already define above. We use the pruning accuracy formula for information gain and error-free accuracy. All the results mention into the table as shown. The information gain accuracy after pruning is up to 90%. We use more than 1.7 million vertices and more than 18.8 million edges to determine the information gain accuracy (Table 5).

According to Figs. 7 and 8 we can observed the fluctuation in objective ranges and the feasibility ratios of dissimilar methods. The optimal solutions obtained by objective values of degree and distance centrality closeness algorithms rise when the value of  $n$  increases because more nodes lead to a greater number of edges.

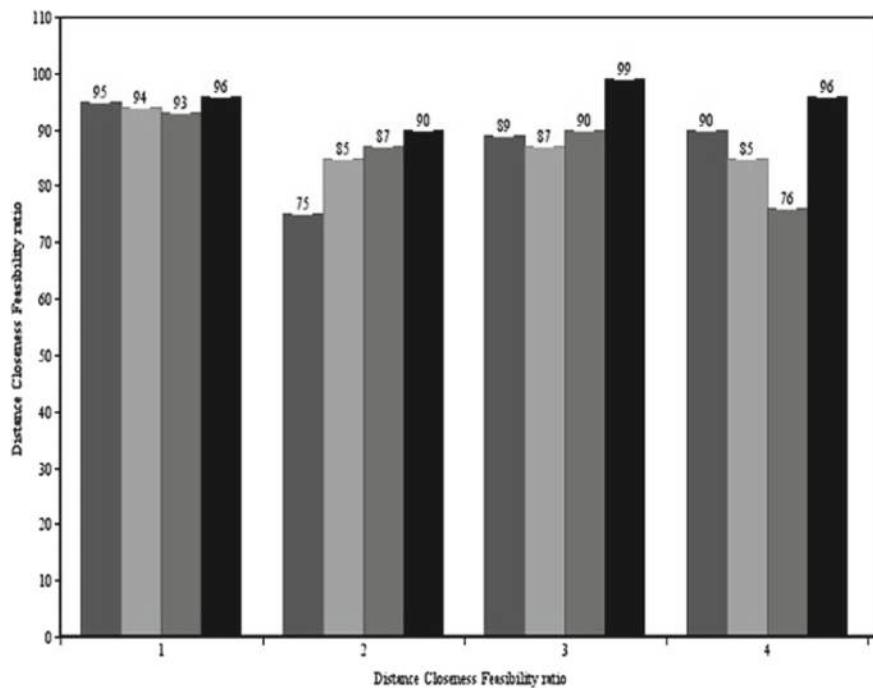


**Fig. 6** Social media 10 years analysis

**Table 5** Accuracy rate with respect to vertices and edges

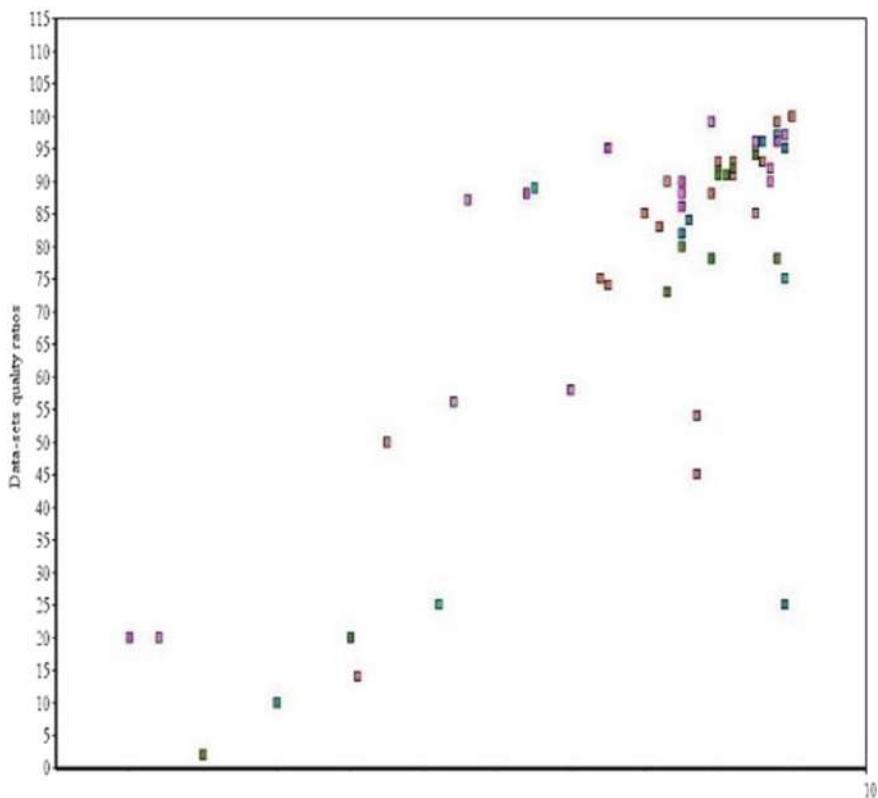
Datasets	Vertices	Edges	Information gain	Accuracy ratio (%)
FB-US	64 k	818 K	15.1	91
FB-L	1.7 m	30 M	13	90
IG-US	46 k	679 K	2.6	95
IG-L	317 k	1.1 M	18.8	97

Conversely, both FB-US and IG-US experience considerably higher objective values, and IG-L and IG-US have indigent feasibility ratios. Owing to the fact that FB-US and IG-US do not well use the information brought by the information gain and information ratio in threshold graphs for every dataset category when the average closeness score is higher, it is likely to have more tenuous users in the community. Moreover, there are many communities with large feasibility scores in IG L that have IO ratios close to 1. This means that high IO score users in IG L are additionally more likely to create homogeneous groups. By default, you see the users who are used a lot in many communities which deceive that FB-US is widely (although never represented) and there appears to be a large amount of IG-L among all other use cases particularly at FB-L as shown at Fig. 3c. But if we had taken a closer look at these communities, we would have discovered that those are the small communities (of size 6 approximately) since direct link users are less likely to be active. In contrast, in IG- L the less prominent IO users are to be found compulsorily as soon as degree closeness scores become larger. This is because they can discover people in Instagram



**Fig. 7** Distance closeness accuracy

more easily (not necessarily by being friends) and view, like or follow these folks on IG.



**Fig. 8** Social media-based addicted community

## 6 Conclusion

In this paper, we make try to attempt to identify social media addiction make us tenuous in real life. We propose some question-based scenarios, asking the people to agree or disagree. We conduct this survey in China and Pakistan. We design two polynomial-time algorithms in which use DADCC. We design pruning-based algorithm, check our results as information gain, information ratio and gain. These algorithms are much efficient and effective to find out the tenuous community. In future, we plan to extend this work, more efficient algorithm design, and another application of life that has a tenuous impact. We have a plan to use blockchain technology with online community detection and also use deep learning with data mining.

## References

1. Shen CY, Huang LH, Yang DN, Shuai HH, Lee WC, Chen MS (2017) On finding socially tenuous groups for online social networks. In: Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining. <https://doi.org/10.1145/3097983.3097995>
2. Jin D et al (2023) A survey of community detection approaches: from statistical modeling to deep learning. *IEEE Trans Knowl Data Eng* 35(2). <https://doi.org/10.1109/TKDE.2021.3104155>
3. Qadir M, Safder MH, Sumra IA (2019) Impact of social media on human life. *Eng Sci Technol Int Res J* 4(1)
4. Guidi B, Michienzi A (2021) Dynamic community structure in online social groups. *Inf (Switz)* 12(3). <https://doi.org/10.3390/info12030113>
5. Wang Y, Cao J, Bu Z, Jiang J, Chen H (2021) Proximity-based group formation game model for community detection in social network. *Knowl Based Syst* 214. <https://doi.org/10.1016/j.knosys.2020.106670>
6. Gaesser B, Keeler K, Young L (2018) Moral imagination: facilitating prosocial decision-making through scene imagery and theory of mind. *Cognition* 171. <https://doi.org/10.1016/j.cognition.2017.11.004>
7. Saleh Al Rasheed L (2022) The effects of a psycholinguistic approach to multisensory instruction on psycholinguistic abilities of children with learning disabilities. *Psycholinguistics* 32(1). <https://doi.org/10.31470/2309-1797-2022-32-1-143-162>
8. Petermann F (2018) Group therapy. *Kindheit und Entwicklung* 27(4). <https://doi.org/10.1026/0942-5403/a000259>
9. Zhao Q, Li L, Chu Y, Yang Z, Wang Z, Shan W (2022) Efficient supervised image clustering based on density division and graph neural networks. *Remote Sens (Basel)* 14(15). <https://doi.org/10.3390/rs14153768>
10. Bastami E, Mahabadi A, Taghizadeh E (2019) A gravitation-based link prediction approach in social networks. *Swarm Evol Comput* 44. <https://doi.org/10.1016/j.swevo.2018.03.001>
11. Li Z et al (2018) Discovering hierarchical subgraphs of K-core-truss. *Data Sci Eng* 3(2). <https://doi.org/10.1007/s41019-018-0068-2>
12. Kumpulainen I, Tatti N (2024) Dense subgraphs induced by edge labels. *Mach Learn* 113(4). <https://doi.org/10.1007/s10994-023-06377-y>
13. Staples L (2012) Community organizing for social justice: grassroots groups for power. *Soc Work Groups* 35(3). <https://doi.org/10.1080/01609513.2012.656233>
14. Moseley C, Bonner E, Ibey M (2021) The impact of guided student-generated questioning on chemistry achievement and self-efficacy of elementary preservice teachers. *Eur J Sci Math Educ* 4(1). <https://doi.org/10.30935/scimath/9448>
15. Shen CY, Yang DN, Huang LH, Lee WC, Chen MS (2016) Socio-spatial group queries for impromptu activity planning. In: IEEE transactions on knowledge and data engineering. <https://doi.org/10.1109/TKDE.2015.2468726>
16. Ji J, Li Z, Xu S, Ge Y, Tan J, Zhang Y (2023) Efficient non-sampling graph neural networks. *Inf (Switz)* 14(8). <https://doi.org/10.3390/info14080424>
17. Shen CY et al (2022) On extracting socially tenuous groups for online social networks with k-triangles. *IEEE Trans Knowl Data Eng* 34(7). <https://doi.org/10.1109/TKDE.2020.3025911>
18. Hönes HC (2021) Gandy's Arkatology: the deluge and romantic climatologies of architecture. <https://doi.org/10.1017/arh.2021.13>
19. Sotirov R (2020) On solving the densest k-subgraph problem on large graphs. *Optim Methods Softw* 35(6). <https://doi.org/10.1080/10556788.2019.1595620>
20. Fuster H, Chamarro A, Oberst U (2017) Fear of missing out, online social networking and mobile phone addiction: a latent profile approach. *Aloma: Revista de Psicología, Ciències de l'Educació i de l'Esport* 35(1). <https://doi.org/10.51698/aloma.2017.35.1.22-30>
21. Bai XQ, Li B, Xu CD, Zhang X (2023) Fast algorithm for the rainbow disconnection coloring of 2-trees. *J Oper Res Soc China*. <https://doi.org/10.1007/s40305-023-00498-w>

22. Berreman G (1963) Caste and community development. *Hum Organ* 22(1). <https://doi.org/10.17730/humo.22.1.kv458786173j14m9>
23. Biswas K, Shivakumara P, Pal U, Lu T (2023) A new ontology-based multimodal classification system for social media images of personality traits. *Sig Image Video Process* 17(2). <https://doi.org/10.1007/s11760-022-02259-3>
24. Hou Y, Xiong D, Jiang T, Song L, Wang Q (2019) Social media addiction: its impact, mediation, and intervention. *Cyberpsychology (Brno)* 13(1). <https://doi.org/10.5817/CP2019-1-4>
25. Darko EM, Kleib M, Olson J (2022) Social media use for research participant recruitment: integrative literature review. <https://doi.org/10.2196/38015>
26. Becker F, Skirzyński J, van Opheusden B, Lieder F (2022) Boosting human decision-making with AI-generated decision aids. *Comput Brain Behav* 5(4). <https://doi.org/10.1007/s42113-022-00149-y>

# Global Path Planning Based on Improved Ant Colony Optimization Algorithm



Ruoyu Li, Jiangwen Deng, and Kaijin Qiu

**Abstract** Ant Colony Optimization (ACO) Algorithm, a probabilistic-based algorithm for seeking optimized paths in graphs, which is originated from the behavior of ants hunting for food. In nature, ants would emit a type of secretion called pheromone when they walk. Other ants may follow the paths where the concentration of pheromone is relatively high, and this would further increase the possibility of choosing the paths. However, traditional ACO algorithm possesses some defect that it might fall into the local optimum. Our strategy is to integrate artificial potential field for traditional ACO algorithm. Consequently, our improved version has prominent improvements with respect to its running time and how the path is chosen. Also, we proposed a new pheromone volatilization mechanism, which efficiently solve the problem of falling into the local optimal and would speed up the convergence of the algorithm.

**Keywords** Ant colony algorithm · Potential field · Global path planning · Triangular pruning

## 1 Introduction

Global path planning refers to look for an obstacle-free and accessible optimum solution from the starting point to destination in a given environment. Researchers have already made tremendous achievements in path planning and innovated various

---

R. Li

School of Communication, Hong Kong Baptist University, Hong Kong, China  
e-mail: [21250294@life.hkbu.edu.hk](mailto:21250294@life.hkbu.edu.hk)

J. Deng

Institute of Deep Learning, Southwest University, Chongqing, China  
e-mail: [djw1579917@swu.edu.cn](mailto:djw1579917@swu.edu.cn)

K. Qiu (✉)

College of Computer and Information Science, Southwest University, Chongqing, China  
e-mail: [qkjswu@163.com](mailto:qkjswu@163.com)

algorithms designate for finding the optimal path, including Dijkstra's algorithm, which people are most familiar with, RRT and PRM [1] algorithms based on random sampling, ACO algorithms based on heuristics, etc.

In this paper, the conventional ant colony algorithm gets improved by combining techniques of potential field and triangular pruning, and the proposed ACO algorithm is capable of getting a better solution and finally accomplishes the desired path. The MATLAB simulation outcomes demonstrate that the proposed algorithm can reduce the running time, as well as the number of iterations of the algorithm and the number of corners of the optimal solution on the premise of ensuring that a passable path is found, so that the UAV is enabled to reach the specified target coordinates in the environment faster and with higher accuracy.

### Contributions

Our contribution is summarized to threefold: (1) We propose a strategy for the traditional ACO algorithm by incorporating artificial potential fields. (2) We propose a new pheromone evaporation mechanism that solves the case that the algorithm might fall into a local optimal solution and speeds up the convergence process.

## 2 Related Works

The following Table 1 provides a detailed comparison of both advantages and disadvantages of some ubiquitous path planning algorithms.

**Table 1** Advantages and disadvantages of some ubiquitous path search algorithm

Algorithm	Advantages	Disadvantages
A* [2]	Heuristic search, simple calculation, fast search speed	Heuristic function is not able to describe the true cost of use accurately
Dijkstra [3]	Simple algorithm, sure to find an optimal path	Long traverse time, low efficiency
Genetic algorithm [4]	Strong robustness, high efficiency	Slow convergence
Artificial bee colony [5]	Strong adaptability	Over dependent on environmental parameter
Simulated annealing [6]	With asymptotic convergence	Long optimization process
RRT [7]	Quick search, No space modeling required	Waste of arithmetic, easy to fall into local minima
Particle swarm optimization [8]	Strong robustness, fast convergence	Easy to fall into local minima
Ant colony [9, 10]	Strong parallelism and global optimization	Easy to fall into local minima, difficult in parameter tuning

Among these the ACO algorithm has been continuously researched and improved to increase its efficiency and practicality in solving various real-world problems, providing an effective method for optimization problem solving.

Dorigo [10] proposed the Ant Colony Algorithm, and the initial application scenario was the Tourist Trader Problem (TSP). It has the advantages of good generality, robustness, algorithmic efficiency, and path smoothing, thereby attracting the attention of major researchers. The improvement directions include: adaptive parameter tuning, local search strategy, parallelization and distribution, ant colony population and strategy optimization, hybrid algorithms.

The initial version of the mathematical modeling of ACO algorithm is primarily embodied in the following three aspects: pheromones release formula, pheromones update mechanism as well as probability transition formula. Pheromones release formula could be expressed by:

$$\Delta \tau_{ij}^k(t, t+1) = \begin{cases} \frac{Q}{L_k}, & k \text{ th ant from } i \text{ to } j \\ 0, & \text{Others} \end{cases} \quad (1)$$

The meaning of the above formula is the pheromones value released by the ant with index  $k$ , when it passes through the point  $j$  from the point  $i$  at the time  $t$ , where  $Q$  stands for pheromone intensity parameter,  $L_k$  denotes the path of the ant with index  $k$  in the current iteration.

A concrete interpretation of the probability transition is, assuming the choice probability of the ant with index  $k$  from the current point  $i$  to the next point  $j$  is determined by  $\tau_{ij}(t), \eta_{ij}(t)$

$$p_{ij}^k(t) = \begin{cases} \frac{[\tau_{ij}(t)]^\alpha * [\eta_{ij}(t)]^\beta}{\sum_s \text{allowed}_k [\tau_{is}(t)]^\alpha * [\eta_{is}(t)]^\beta}, & s \in \text{allowed}_k \\ 0, & s \notin \text{allowed}_k \end{cases} \quad (2)$$

where  $p_{ij}^k(t)$  is the probability that at the time  $t$ , the ant with index  $k$  selects  $j$  as its next point when it is currently at the point  $i$ , and  $\tau_{ij}(t)$  is the pheromone concentration of the point  $i$  and  $j$  at the time  $t$ .  $\eta_{ij}(t)$  denotes the distance between  $i$  and  $j$  at the time  $t$ .  $s$  denotes to the next accessible point. Similarly,  $\tau_{is}$  denotes to the pheromone concentration of  $i$  and  $s$ .  $\eta_{is}$  denotes to the distance between  $i$  and  $s$ .  $\alpha$  denotes the pheromone influence factor while  $\beta$  denotes the heuristic factor.  $\text{allowed}_k$  denotes the set of points that the current point  $i$  could choose from as the next point. Regarding the value of  $\eta_{ij}$ , assume the coordinate of  $i$  is  $(x_i, y_i)$ , the coordinate of  $j$  is  $(x_j, y_j)$ , then the value of  $\eta_{ij}$  is the inverse of the Euclidean distance between points  $i$  and  $j$ , which is expressed as follows:

$$\eta_{ij}(t) = \frac{1}{D_{ij}} D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (3)$$

From the above formula, it can be inferred that with the increasement of the Euclidean distance between point  $i$  and point  $j$ , the smaller the heuristic distance  $\eta_{ij}$  between them will be.

Pheromone update mechanism is vital in ACO algorithm. In order to avoid falling into the local minima solution due to the excessive pheromone concentration of a certain path. The pheromone released on the current path by ants each time needs reasonably evaporated, and the pheromone evaporation and update mechanism is expressed below:

$$\begin{aligned}\tau_{ij}(t+1) &= (1 - \rho)\tau_{ij}(t) + \Delta_{ij}(t) \\ \Delta_{ij}(t) &= \sum_{(k=1)}^m \Delta\tau_{ij}^k(t)\end{aligned}\quad (4)$$

where  $\rho$  denotes pheromone evaporation coefficient, which is in the range of  $[0,1]$ ,  $\Delta\tau_{ij}$  the increase of pheromone at moment  $t$ , that is, the volume of pheromone released by ant on the current path, which can be obtained by formula (1). As shown in the Fig. 1, it illustrates the flow chart of the conventional ACO algorithm.

### 3 Methods

#### 3.1 Improvements of Heuristic Function

The artificial potential field [11] is derived from classical mechanics, and in the algorithm, it specifically analyzes the force relationship of a robot. The target node will produce gravitational force on the robot, while the obstacles in the path will inversely provide repulsive force in different directions, abstracting the surroundings as a potential field, where the target node guides the robot forward and the obstacles repel the robot's movement. The definition of the conventional gravitational potential field can be derived from the following formula:

$$U_{ap}(X_c) = \frac{1}{2}K_a \cdot d^2(X_c, X_g) \quad (5)$$

where  $X_c$  denotes the current coordinate value of the robot in the global map,  $X_g$  denotes the coordinate of the target point, and  $K_a$  denotes a gravitational potential field constant while  $d^2(X_c, X_g)$  denotes the Euclidean distance from the current point  $g$  to the target point. The definition of the conventional gravitational potential field can be derived from the following formula:

$$U_{rep}(X_c) = \begin{cases} \frac{1}{2}K_{rep} \left( \frac{1}{d(X_c, X_o)} - \frac{1}{d_m} \right)^2, & d < d_m \\ 0, & d \geq d_m \end{cases} \quad (6)$$

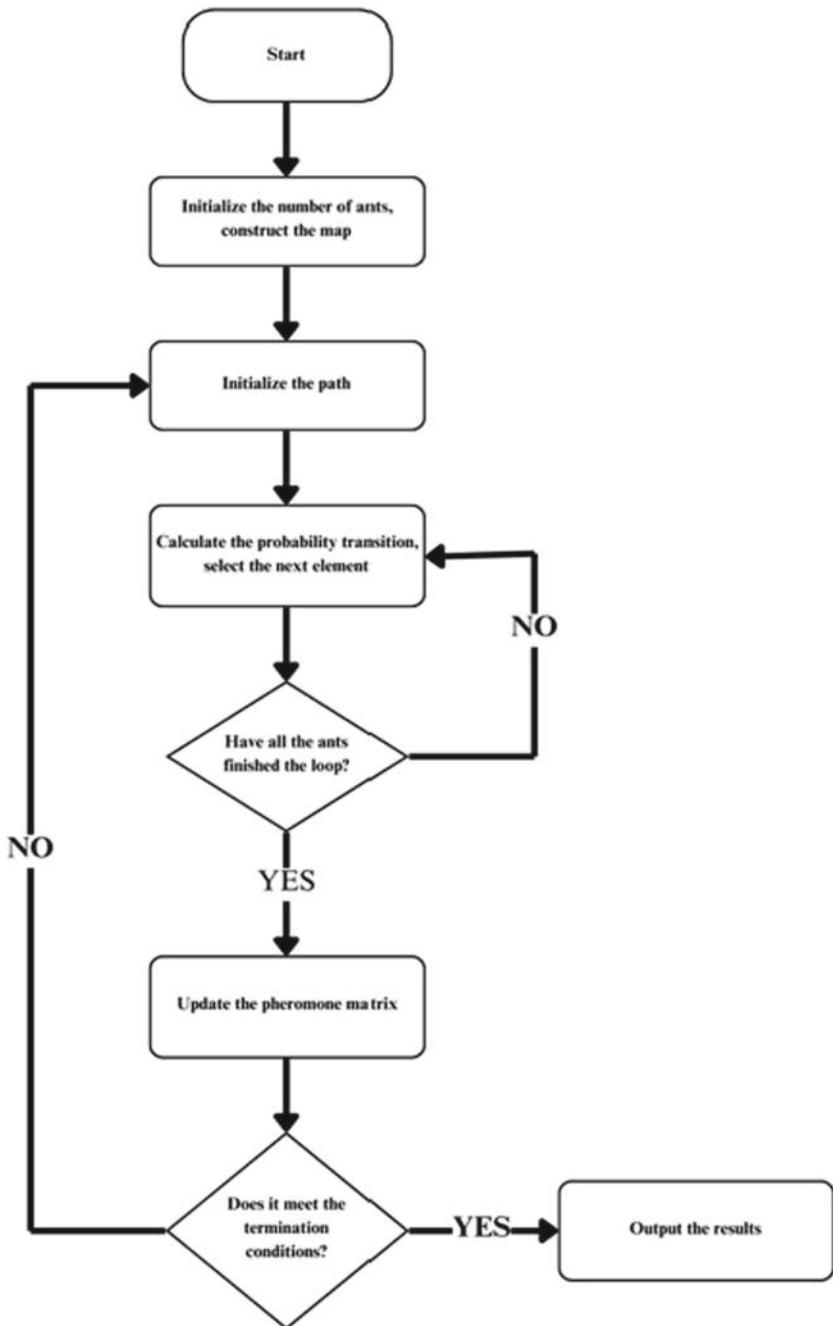


Fig. 1 Flow chart of the traditional ant colony algorithm

Similar to the formula (5),  $K_{\text{rep}}$  is the repulsive potential field constant,  $X_o$  denotes the coordinate of obstacles in the global map,  $d(X_c, X_o)$  denotes the Euclidean distance between the present point  $c$  and the obstacle  $g$ , and  $d_m$  refers to the safe distance of the repulsive potential field. It can be seen that within the safe range, the further the robot and the obstacles are, the smaller the repulsive force between them would be, and vice versa. The global potential field function is obtained according to the force analysis as follows:

$$U(X_c) = \begin{cases} \frac{1}{2}K_a \cdot d^2(X_c, X_g) + \frac{1}{2}K_{\text{rep}}\left(\frac{1}{d(X_c, X_o)} - \frac{1}{d_m}\right)^2, & d(X_c, X_g) < d_m \\ \frac{1}{2}K_a \cdot d^2(X_c, X_g), & d(X_c, X_g) \geq d_m \end{cases} \quad (7)$$

Formula (8) indicates that in the traditional ACO algorithm, the heuristic information  $\eta_{ij}(t)$  is only related to the distance between point  $i$  and point  $j$ . With the increasement of the Euclidean distance, the smaller the value of the heuristic information is. Since the numerator is fixed, which make it difficult for the ACO algorithm to converge, we propose a heuristic function that incorporates the dynamic weight change of the artificial potential field.

$$\eta_{ij}(t) = \frac{1}{D_{ij}} \quad (8)$$

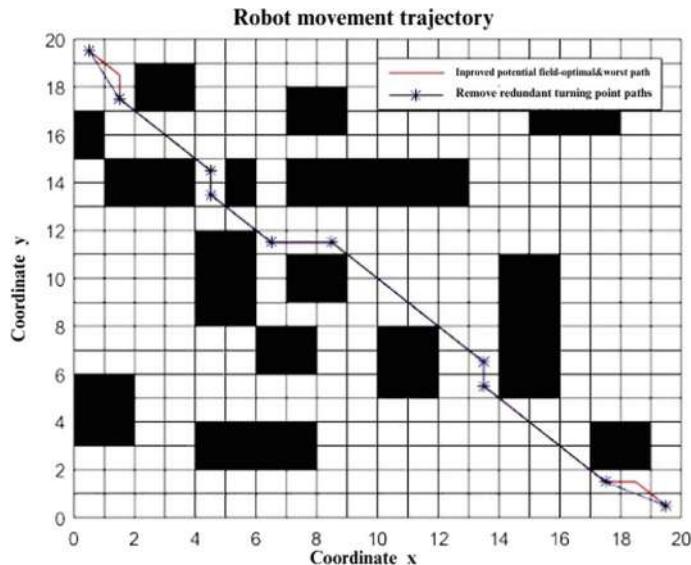
$$\eta_{ij}(t) = \frac{0.5(\text{sigma} \cdot D_{ij} \cdot U(X_c) \cdot \cos\theta)}{D_{ij}} \quad (9)$$

where sigma is the potential field constant greater than zero,  $D_{ij}$  refers to the Euclidean distance between point  $i$  and point  $j$ ,  $U(X_c)$  here represents the resultant force of the gravitational force and the repulsive force, which is directed and  $\theta$  denotes the angle between the path and the resultant force.

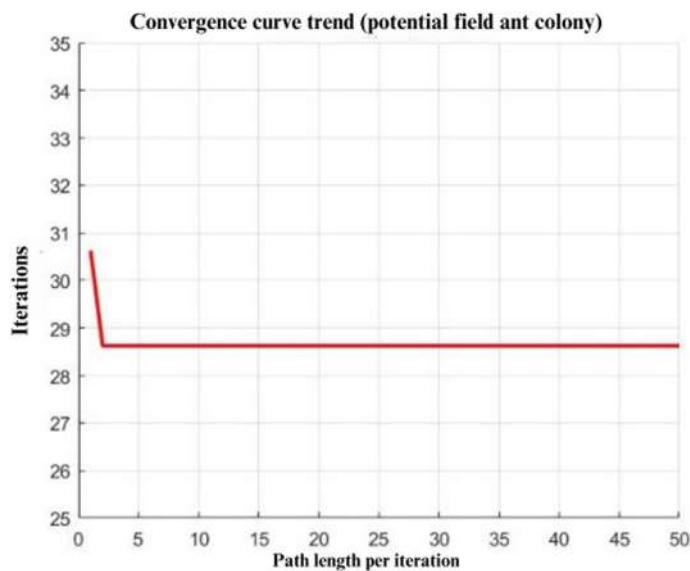
The following figure shows the result graph of MATLAB simulation. Figure 2 is the global motion trajectory of the robot with the fused artificial potential field, Fig. 3 is the iterative convergence trend of the ACO algorithm with the fused artificial potential field, and Fig. 4 is all the paths of the ants in the colony.

### 3.2 Improvements of Pheromone Evaporation Strategies

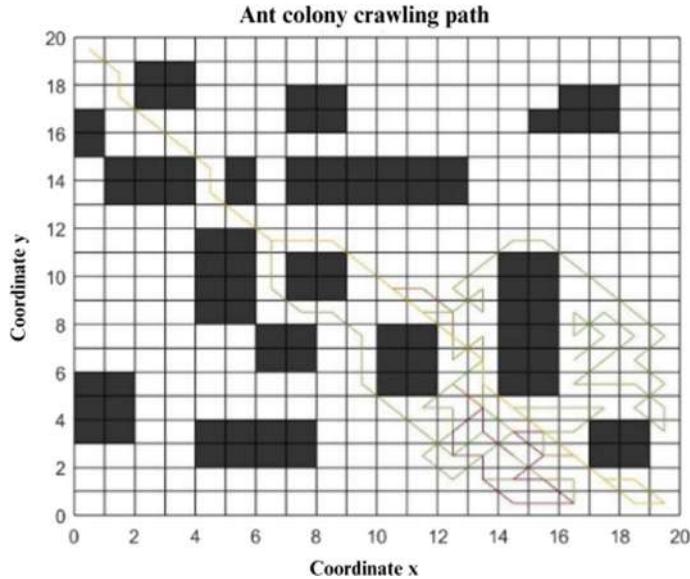
Because the ants behind will be affected by the pheromones released by the previous ones. If the previous ants choose the same path very often, they would leave excessive pheromone on the path, exceeding the predetermined threshold, which may cause the ants behind to choose the path blindly, leading the algorithm to result in a local optimum, which is very unfriendly to the convergence of the algorithm. To address this issue, we proposed a new pheromone release algorithm, on the basis on the conventional algorithm, we added the differences between the optimal path and the



**Fig. 2** Global path of the artificial potential field



**Fig. 3** Convergence trend of the artificial potential field



**Fig. 4** Ant colony crawling path fused with artificial potential field

worst path. Thus, the algorithm will evaporate more pheromone on the path, so that the algorithm slows down to the local optimal solution situation.

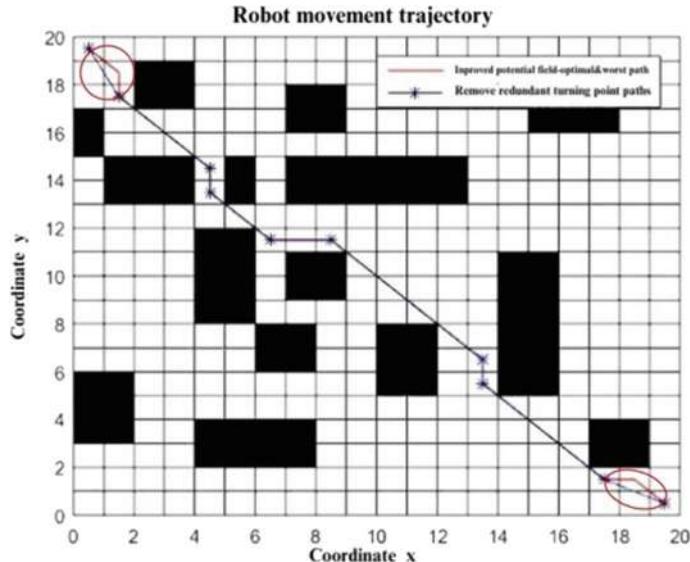
$$\Delta \tau_{ij}^k(t, t+1) = \begin{cases} \frac{Q}{L_k} + \frac{bQ}{\min L_k} - \frac{wQ}{\max L_k}, & \text{kth ant from i to j} \\ x, & \text{Others} \end{cases} \quad (10)$$

### 3.3 Path Pruning Optimization

Considering the existence of time cost and energy loss problems in global path planning, the final planned path should realize the goals of small number of inflection points, high path smoothness, and short actual displacement of the robot. In this paper, we proposed a triangular pruning [12] strategy to cut off the redundant inflection points to reduce the time cost and energy loss of the robot, and guarantee the robot will accomplish the desired path.

The specific implementation process can be summarized as follows:

- Step 1: Among the  $n$  inflection points in the global path planning, determine whether there is a straight-line pathway between the inflection point  $n_0$  and  $n_2$ , that is, determine whether there is an obstacle blocking between the inflection points  $n_0$  and  $n_2$ .



**Fig. 5** Results of triangular pruning

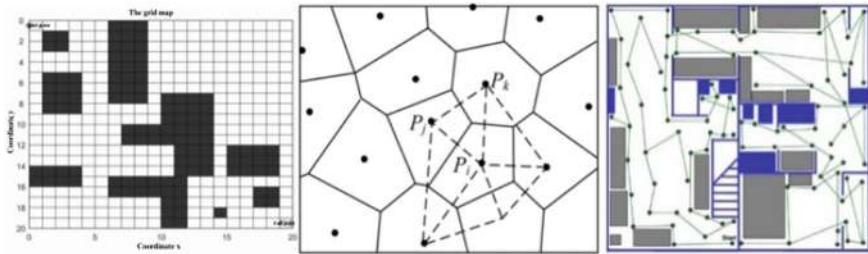
- Step 2: If there is a straight-line path between the two, remove the intermediate inflection point  $n_1$ , connect the inflection points  $n_0$  and  $n_2$ , and perform the first step, or else jump out of the pruning loop for the  $n_0$  inflection point.
- Step 3: Loop the operations of Step 1 and Step 2 to complete the pathway judgment of  $n$  inflection points and complete the pruning operation between neighboring nodes.

The circle in Fig. 5 is the result of triangular pruning. We can see that the triangular pruning strategy allows us to remove redundant inflection points and reduce the distance and energy loss of the robot, and guarantee the robot will accomplish the desired path.

## 4 Results and Discussion

### 4.1 Construction of the Grid Map

The foundation for a robot to accomplish path planning is to have global map information, which includes the location of obstacles, the current coordinates of itself, and the coordinates of the target point. In the past, researchers have put forward different ideas for map construction, including the grid method [13], Voronoi method [14], topology method [15], etc. Among them, the grid method is the simplest one to be



**Fig. 6** Map display in three different formats

employed, with number zero representing a blank pathway and number one representing an obstacle. The construction of the map can be completed by assigning different values to different nodes in the map.

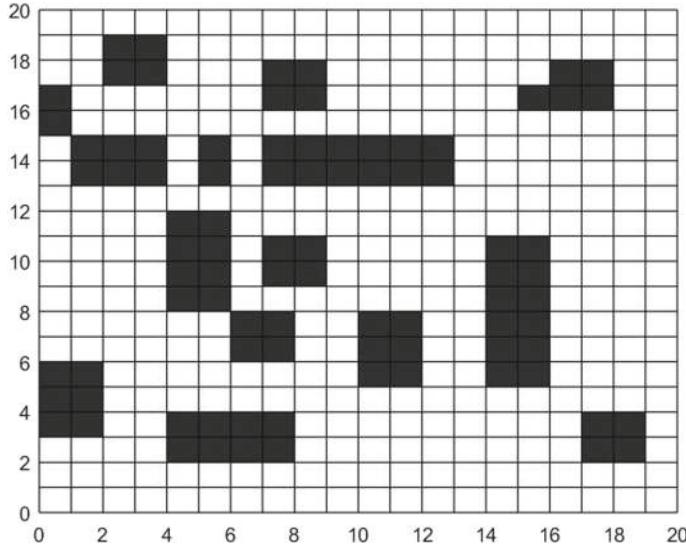
The overall idea of the Voronoi diagram method is to divide obstacles in the map, each polygonal region represents an obstacle. The advantage of this kind of composition is that it can effectively avoid the obstacle collision in the path planning, however, it is not applicable in the dynamic circumstances.

Topological map is a map composed of nodes and edges; it only considers how well the connectivity is between nodes but does not consider the cost of connectivity. Thus, topological maps do not represent well the environmental situation when the structure is complex due to the absence of the details. Figure 6 shows, from left to right, a grid map, a Voronoi map, and a topological map.

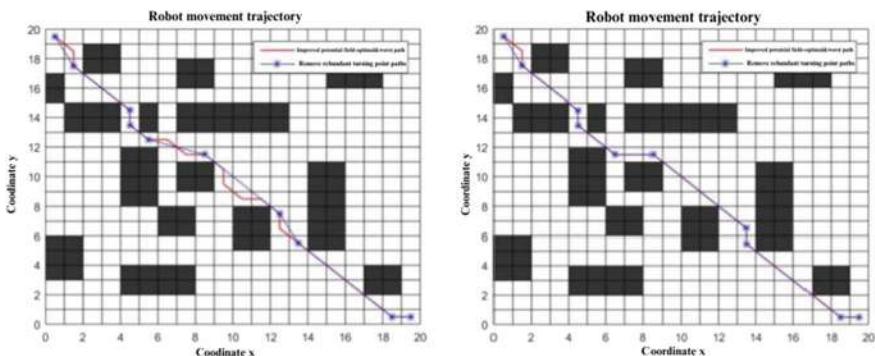
The core idea of the grid method is to characterize the environment in terms of a matrix, where all nodes in the grid map have their own values. The value zero represents the free space, which is represented by a white grid, and value one represents the obstacles, which are represented by a black grid. By setting the percentage of the two parts according to the needs, we can design the grid map we need. Since grid maps have the advantages of simple structure, easy expression, and small amount of data, in this paper, we will carry out the construction of grid maps in MATLAB and complete the basic simulation of the ACO algorithm that incorporates the artificial potential field. The following Fig. 7 shows the required grid map that we have built:

## 4.2 Comparative Analysis of Heuristic Functions

The first of the following Figs. 8, 9, and 10 shows the original heuristic function and the second shows the improved heuristic function (Table 2).



**Fig. 7** Grid map display

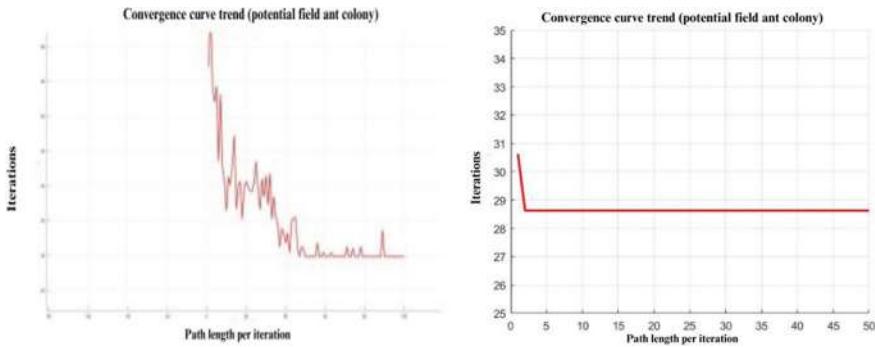


**Fig. 8** Trajectory of global path planning

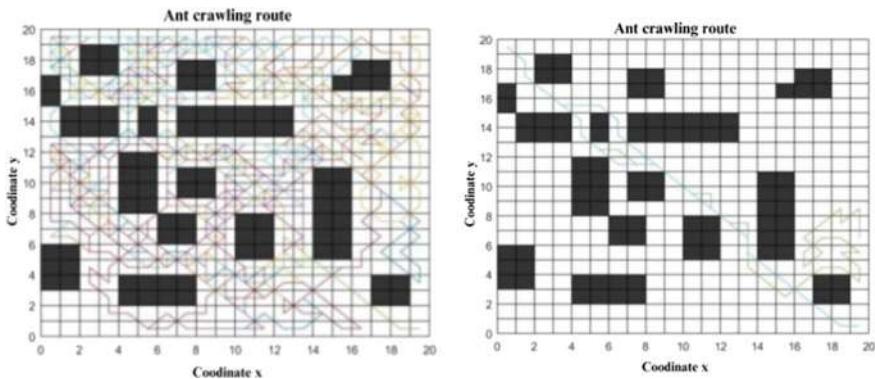
#### 4.3 Comparative Analysis of Pheromone Evaporation Strategies

The first Fig. 11 shows the convergence of the original algorithm, and the second figure shows the convergence of the innovative algorithm:

Through the various comparisons in this chapter, the ACO fused with the artificial potential field has a prominent improvement effect, both ensuring algorithmic running time and convergence performance. The most obvious is that the running time is reduced by about 30%, which is significant for real-time global path planning. The excellent convergence performance of the algorithm can also effectively avoid



**Fig. 9** Convergence trend of the algorithm

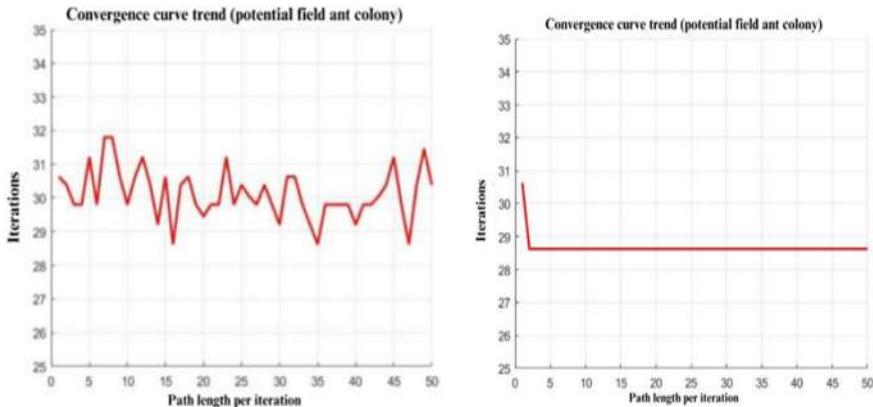


**Fig. 10** Crawling route of an ant colony in the algorithm

**Table 2** Comparison of ant colony algorithms of different heuristics

Comparative item	Original heuristic function	Heuristic functions fused with artificial potential fields
Running time	9.821845	6.249884
Optimal path length	33.7990	28.6274
Remove excess inflection length	28.4268	27.4493

falling into the traditional local optimal solution, which is ineffable for finding the optimal path.



**Fig. 11** Effect of pheromone volatilization strategy on algorithm convergence

## 5 Conclusion

In this paper, we firstly, compare the advantages and flaws of the existing global path searching algorithms, and analyze the specific theory and mathematical model of the traditional algorithm, and put forward our improved strategy of fused artificial potential field for the traditional ACO algorithm, which indeed shows certain improvements compared with the traditional ACO algorithm with respect to the running time and the optimal path of the ACO algorithm.

Also, we propose a new pheromone evaporation mechanism for the conventional ACO algorithm which might result in a local optimum and affects the convergence of the algorithm. As a consequence, it effectively solves the situation where it converges to a local optimum and accelerates the convergence of the algorithm. Then, we compare our ACO algorithm incorporating artificial potential field with the traditional ACO algorithm with respect to path planning, algorithm running time, algorithm convergence, and ACO path exploration through simulation experiments in MATLAB2015. b. Finally, we confirm the feasibility of our algorithm through the experimental results.

## References

1. Li Q, Xu Y, Bu S, Yang J (2022) Smart vehicle path planning based on modified PRM algorithm. Sensors 22(17):6581
2. Liu L, Wang B, Xu H (2022) Research on path-planning algorithm integrating optimization A-star algorithm and artificial potential field method. Electronics 11(22):3660
3. Szczepanski R, Tarczewski T (2021) Global path planning for mobile robot based on artificial bee colony and Dijkstra's algorithms. In: 2021 IEEE 19th International power electronics and motion control conference (PEMC). IEEE, pp 724–730

4. Papazoglou G, Biskas P (2023) Review and comparison of genetic algorithm and particle swarm optimization in the optimal power flow problem. *Energies* 16(3):1152
5. Karaboga D, Basturk B (2008) On the performance of artificial bee colony (ABC) algorithm. *Appl Soft Comput* 8(1):687–697
6. Du Z, Wang Z, Wang Z, Qin W, Duan Y (2011) Improved polymorphic ant colony algorithm with double simulated annealing. *J Cent South Univ (Science and Technology)* 42(10):3112–3117
7. Kuffner JJ, Lavalle SM (2000) RRT-connect: an efficient approach to single-query path planning. In: Proceeding of the IEEE International conference on robotics and automation 2000 proceeding ICRA
8. Wang D, Tan D, Liu L (2018) Particle swarm optimization algorithm: an overview. *Soft Comput* 22:387–408
9. Wu L, Huang X, Cui J, Liu C, Xiao W (2023) Modified adaptive ant colony optimization algorithm and its application for solving path planning of mobile robot. *Expert Syst Appl* 215:119410
10. Dorigo M (2007) Ant colony optimization. *Scholarpedia* 2(3):1461
11. Zhang W, Xu G, Song Y, Wang Y (2023) An obstacle avoidance strategy for complex obstacles based on artificial potential field method. *J Field Robot* 40(5):1231–1244
12. Harabor D, Grastien A (2011) Online graph pruning for pathfinding on grid maps. In: Proceedings of the AAAI conference on artificial intelligence, vol 25(1). pp 1114–1119
13. Chunshu L, Haifeng L, Genqun C (2009) The improved potential grid method in robot path planning
14. Bhattacharya P, Gavrilova ML (2008) Roadmap-based path planning-using the voronoi diagram for a clearance-based shortest path. *IEEE Robot Autom Mag* 15(2):58–66
15. Li B, Liu H, Su W (2019) Topology optimization techniques for mobile robot path planning. *Appl Soft Comput* 78:528–544

# Advanced Automated System for Optimal Management in Public Health Emergencies



**Shakil Muhammad, Adnan Mujahid Khan, Azka Qureshi, Rajakumar Arul,  
and Kalaipriyan ThirugnanaSambandam**

**Abstract** The purpose of this study work is to provide a sophisticated automated system that has been created for effective management in situations that need little human contact during times of public health crisis. A suite of computer vision modules, such as gesture recognition and gender recognition, are included to the system. Each of these modules has been painstakingly customized to handle unique issues that are faced in real-time dynamic contexts. The gesture recognition module makes use of both hand and upper body key points to handle the delicate nuances of a wide variety of human movements within dynamic environments. It does this by deftly reducing obstacles associated to varying lighting conditions, picture quality, and occlusions. This automated system has practical uses outside of venues such as restaurants, amusement parks, and service areas, in addition to the technological achievements it has made. The deployment of this system makes it possible to gather input in real time in an independent and objective manner, which represents a big step forward in the direction of improving public health outcomes.

**Keywords** Automated feedback collection · Computational vision · Dynamic gesture recognition · Gender identification · Real-time interaction framework

---

S. Muhammad

Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea  
e-mail: [shakil@genesislab.com](mailto:shakil@genesislab.com)

A. Mujahid Khan

History Ltd, Coventry, UK  
e-mail: [adnan@coinchance.io](mailto:adnan@coinchance.io); [adnan@histofy.ai](mailto:adnan@histofy.ai)

A. Qureshi

Applab Qatar, Doha, Qatar  
e-mail: [azka@coinchance.io](mailto:azka@coinchance.io); [Azka.q@applab.qa](mailto:Azka.q@applab.qa)

R. Arul (✉) · K. ThirugnanaSambandam

Centre for Smart Grid Technologies, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai Campus, Chennai, India  
e-mail: [rajakumararul@ieee.org](mailto:rajakumararul@ieee.org)

K. ThirugnanaSambandam

e-mail: [kalaipriyan.t@vit.ac.in](mailto:kalaipriyan.t@vit.ac.in)

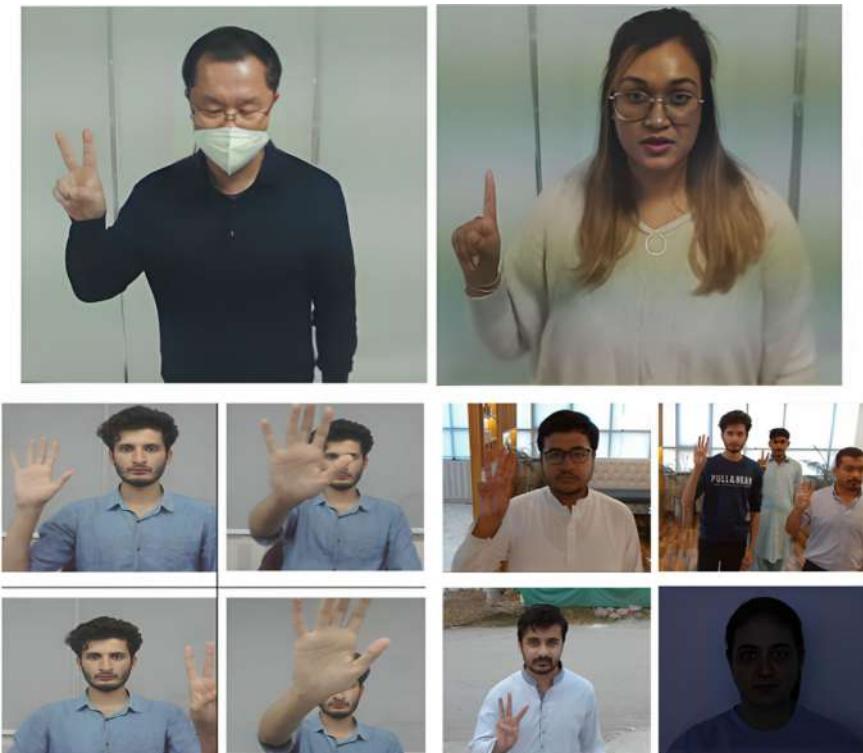
## 1 Introduction

In recent years, deep learning has become a popular and effective technique in the field of computer vision for identifying demographic attributes of humans such as gender, gesture, and face. In this paper, gesture recognition, face detection, and gender classification methods are combined to address three goals: (1) multi-person gesture recognition in indoor and outdoor settings and (2) gender classification for every identified subject; a The corresponding research spans to highlight and address the challenges in gender, and gesture recognition for the proposed real-time application.

Accurate gesture recognition is a difficult task due to unanticipated factors such as environmental noise and language variability. The human hand is exceptionally versatile, capable of forming various static and dynamic gestures. Recognizing similar gestures imposes higher demands on the algorithm. Additionally, when individuals make diverse gestures, it introduces challenges such as occlusion of key points on the hand. This occlusion becomes more pronounced when executing two-handed gestures, like clasping fists, leading to increased difficulty in the algorithm's feature extraction and the prediction of invisible points. In real-world scenarios with a diverse audience exhibiting unique behaviors and poses there is another challenge including the potential for misinterpretation of gestures. Individuals engaged in activities unrelated to providing feedback, such as holding a phone or participating in a conversation, present difficulties for existing hand gesture models to yield desired results in identifying those intending to provide feedback. Addressing these challenges requires strategic approaches to ensure the effectiveness of the proposed algorithm in diverse and real-world scenarios. To overcome these challenges, the authors of this article opted for key point-based approaches instead of relying solely on hand-cropped images. Techniques such as Mediapipe and methodologies discussed in the works of Zhang et al. [1], Oudah et al. [2], Kopuklu et al. [3], Schlusener and Bucker [4], Mahmud et al. [5] were employed.

Most face detection algorithms are designed in the software field and have a high detection rate, but they often need several seconds to detect faces in a single image, a processing speed that is insufficient for real-time applications. Problems associated with state-of-the-art models, revealed issues associated with inaccurate embeddings generated by various pipelines as outlined in Deng et al. [6], Guo et al. [7], and Anwar and Raychowdhury [8]. To address this, the authors opted for the implementation of the InsightFace model, leveraging a ResNet50 backbone developed by An et al. [6], renowned for its superior accuracy. The researchers developed a facial recognition model that is independent of lightning, background, image quality, and image noise, ensuring high performance across various races and genders. This endeavor required a large-scale dataset, which the authors meticulously curated and annotated. Figure 1 shows the sample images from the dataset illustrating gesture and gender detection scenarios.

Concretely, the proposed system tries to solve a threefold objective: firstly, the multi-person gesture recognition in both indoor and outdoor environments; secondly, gender classification for each identified subject. To achieve these goals, a systematic



**Fig. 1** Sample images from dataset illustrating gesture and gender detection scenarios

approach is essential. For extracting gestures in a live video feed, person detection and tracking become imperative. Likewise, gender classification necessitates face detection to extract relevant facial features. Additionally, to circumvent the challenge of the same subject being repeatedly accounted for, especially in the context of mask-wearing, it becomes crucial to integrate mask detection and face recognition technologies into the proposed system. An in-depth examination of each of these technologies is provided to the reader in the next part, which also serves to provide a brief summary of the background information. In the sophisticated artificial intelligence-driven system that has been presented, this foundation is necessary for understanding the delicate interaction of person identification, tracking, gesture recognition, face detection and recognition, as well as mask detection.

## 2 Related Works

In the scientific literature on person identification, there has been a profusion of different techniques, each of which aims to improve accuracy and efficiency in a variety of different settings. The introduction of deep learning marked the beginning of a new age. Notable milestones include SSD, which was developed by Liu et al. in 2016 [9] and Faster R-CNN, which was developed by Ren et al. in 2017 [10]. Both of these networks include regional proposal networks in order to increase the accuracy of object detection. By splitting the picture into a grid and concurrently predicting bounding boxes and class probabilities for each grid cell, the You Only Look Once (YOLO) framework that was developed by Redmon et al. in 2016 [11] stood out as a revolutionary framework. It was the first to pioneer a unified method.

YOLO is an important option for real-time person recognition applications in dynamic environment. YOLOv7, developed by Wang et al. in the year 2023 [12], displays greater performance in comparison with existing methods, shows high performance in terms of both speed and accuracy measures. Its operating method suggests the improvements in object identification are used to have a favorable influence on gesture identification. YOLOv7 presents a novel technique by including a joint detection head, which makes it possible to make predictions of object bounding boxes and skeleton key points simultaneously. The traditional approaches to gender categorization often depended on characteristics that were produced by hand and machine learning algorithms. Tan and Triggs [13] used Local Binary Pattern descriptors in conjunction with support vector machines (SVMs) in order to classify individuals according to their gender. On the other hand, convolutional neural networks (CNNs) have revolutionized gender categorization with abstract feature extraction, by Goodfellow et al. in 2016 [14].

The monitoring of individuals is a primary focus of study in the field of computer vision. Simple Online and Real-Time Tracking (SORT) [15] algorithm utilizes a hybrid version with Kalman filtering and the Hungarian algorithm in order to perform real-time tracking. Authors Wojke and Bewley [16], developed Deep Simple Online and Real-Time Tracking (DeepSORT), a technology that enhances the capabilities of SORT by including deep appearance traits. This enhances the tracking accuracy in situations that include occlusions and identity shifts. Additionally, recent developments have led to the development of the Bounding Box Transformer for SORT (BoTSort) algorithm. This approach involves the application of transformer-based structures to the problem of person tracking [17].

Hand gestures serve as a nonverbal means of communication with applications spanning many sectors, including communication for deaf-mute persons, robot control, home automation, and medical situations. Several research works make use of a wide range of methodologies, including computer vision and sensor technologies. A comprehensive study conducted by [18] to examine a variety of gesture recognition systems, with a particular emphasis on the extraction of features and classification algorithms. On the other hand, feature-based techniques [19] have obstacles that are associated with constraints in the scene backdrop, different lighting circumstances,

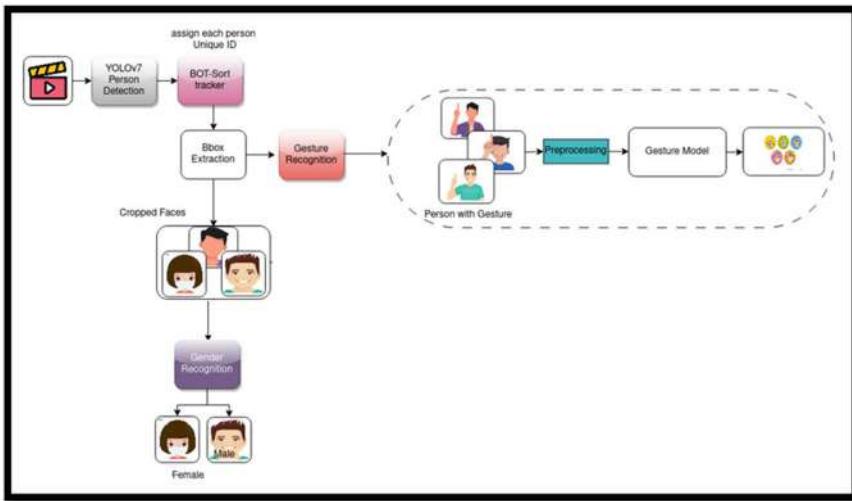
algorithmic accuracy in feature extraction, dataset characteristics, the selection of classification algorithms, and application specialization. Although one alternative that has gained widespread acceptance involves extracting key points directly from the hand for use in classification systems [20], the effectiveness of this approach is hindered by challenges that arise from a variety of backgrounds and hand orientations, which in turn affects the accuracy of the system. However, an effective architecture for gender identification and gesture recognition is still in debate.

### 3 System Architecture

To allow thorough human interaction analysis, the system that has been presented incorporates two separate procedures. The first phase of the Real-Time System is characterized by its dynamic operation, which is accomplished by continually recording frames in real time. The simultaneous identification of persons inside each frame, the subsequent monitoring of recognized individuals, the recognition of motions if they are there, and the collection of this information into an array are all tasks that it is charged with. The system makes use of a consensus process applied over several frames in order to guarantee the robustness of gesture identification. This allows it to overcome the constraints that are connected with snapshot-based analyses. On the other hand, Real-Time System takes snapshots consisting of prospective face areas, so saving them for further analysis.

Complex tasks comprised of the backend system, with the conjunction of Real-Time System. The array of face pictures is received, and begins the process of face detection. The system produces feature embeddings, made it easier to do effective feature matching in order to determine whether the detected face is present within the existing data. On such existence the subsequent analysis is not carried out, and there is no measurement that is recorded. The face is not included in the records, a gender categorization is carried out, and the findings of the survey are stored in a relational database. This two-fold system design enables an accurate study of human interactions by combining real-time dynamics with backend processing for a holistic knowledge of the observed events (Fig. 2).

In this research work, the use of the capabilities of face embeddings that were generated by the InsightFace is imposed. ResNet50 base model is used to guarantee the robustness and invariance of the proposed gender detector. These conditions include variations in lighting, background, races, facial color, picture noise, masks, and other occlusions. The face characteristics were represented both compact and comprehensive by these embeddings that provides wealth of information for gender categorization. To determine the gender of a person, the authors used a straightforward SVM approach. Utilizing the high-dimensional representation of the InsightFace embeddings, the SVM was trained on the pre-trained features derived from the embeddings. During the training phase, the previously taught features of the InsightFace embeddings were fed into the SVM, from where the fundamental patterns and boundaries that differentiate between the various genders. The trained SVM model



**Fig. 2** System workflow

was used as the suggested gender detector, which was able to properly estimate the gender of people based on the attributes that were extracted.

## 4 Experimental Analysis

The public datasets were integrated into the assessment process, primarily for the purpose of evaluating the performance of the modules that were responsible for gesture recognition and gender evaluation. As part of the planned research, two Korean datasets were used, each of which included 75,785 and 158,400 face photos, respectively. The authors further integrated the All-Age-Faces dataset and the FEI dataset to the Korean datasets to generate a dataset, that the authors call mix-face dataset.

In computer vision, specifically in the context of assessing the performance of real-time tracking systems, metrics such as Multiple Object Tracking Accuracy (MOTA) and Higher Order Tracking Accuracy (HOTA) are commonly used. Both MOTA and HOTA are comprehensive metrics designed to evaluate the overall performance of object tracking systems. It considers multiple factors, including false positives, false negatives, and identity switches, providing a holistic assessment of the tracking accuracy.

The Real-Time System's performance evaluation focused on two key aspects: (1) the accuracy of the detection and tracking system Tables 1 and 2 the accuracy of the gesture recognition system (Table 2). Beyond assessing the overall system performance, the authors conducted a thorough examination of individual components

to ensure their efficiency and accuracy. This detailed evaluation not only validated the system's end-to-end performance but also pinpointed potential bottlenecks and refined specific aspects.

For the detection and tracking evaluation, as discussed earlier, the authors utilized the HOTA matrix. Performance of the proposed detection and tracking system across diverse scenarios, emphasizing both accuracy and reliability. On the Indoor Single Person, Outdoor Single Person, Indoor Multi Person, and Outdoor Multi-person scenarios, as well as the Combined outcomes, which are reported in Table 1. Within the Indoor Single Person scenario, included with a total of ten movies and more than nine thousand frames, the system was able to effectively identify 104 distinct IDs with an impressive detection accuracy and tracking accuracy of 99% for each, respectively. In the Outdoor Single person scenario, which consisted of 14 films and 10,000 frames, the system displayed performance capabilities that were comparable. Hybridized detection and tracking, enabled to identify 200 distinct IDs with a 99% accuracy rate overall. The technology shown impressive performance in situations that included numerous individuals and took place both inside and outdoors. For the "Indoor Multi Person" situations (15 movies, 5800 + frames), the system was able to successfully monitor 506 distinct IDs, whereas for the "Outdoor Multi Person" scenarios (13 films, 4800 + frames), it was able to trace 466 unique IDs. The system

**Table 1** Performance metrics for object detection and tracking in various indoor and outdoor scenarios

Scenario	No. of Videos/frames	No. of unique IDs	No. of dets	Detection acc	Tracking acc
Indoor single	10/9021	104	17,818	0.99	0.99
Outdoor single	14/10000	200	20,288	0.99	0.99
Indoor multiple	15/5876	506	19,913	0.98	0.98
Outdoor multiple	13/4857	466	18,811	0.98	0.98
Combined	52/29754	1276	76,830	0.985	0.985

**Table 2** Comparison of gesture recognition

Scenario	No. of Videos/frames	Gesture accuracy (Hand-only key points)	Gesture accuracy (Hand & upper limb key points)
Indoor single	10/9021	0.784	0.953
Outdoor single	14/10000	0.714	0.964
Indoor multiple	15/5876	0.704	0.925
Outdoor multiple	13/4857	0.684	0.896
Combined	52/29754	0.720	0.930

was able to handle complicated situations with several entities, as shown by the fact that the identification and tracking accuracies were maintained at a strong 98% for both scenarios. The results of the single person and multi-person situations, as well as the indoor and outdoor settings, are combined, it is shown that the system successfully tracked more than 1250 distinct people, with an overall detection and tracking accuracy of 98.5%. This was accomplished by using a total of 52 movies covering around 30,000 frames. The system is capable to adapt to a wide variety of situations, which makes it a dependable option for applications in the real world where precise identification and tracking are essential.

An in-depth analysis of the suggested gesture recognition system is shown in Table 2, which compares it to an existing model developed by Guerrieri and Parla. Each of the situations that were evaluated consisted of more than fifty movies and roughly thirty thousand frames. These scenarios included the Indoor Single scenario, the Outdoor Single scenario, the Indoor Multiple scenarios, and the Combined scenario. The existing model, which is based on hand-only key points and was developed by Guerrieri and Parla, displays variable accuracy ratings depending on the circumstance. Specifically, Indoor Single Person scenarios achieved an accuracy of 0.784, while Outdoor Single Person scenarios had a slightly lower accuracy of 0.714. For Indoor and Outdoor Multi-person scenarios, the accuracies were 0.704 and 0.684, respectively. The Combined scenario yielded an overall accuracy of 0.72. The observed decrease in accuracy for outdoor scenarios suggests challenges in handling environmental factors affecting hand-only key points detection in outdoor settings. In contrast, the proposed model integrates both Hands and Upper Body key points, resulting in significantly enhanced accuracy across all scenarios. This contextual information contributes to the improved accuracy observed, with the model achieving 0.953 in the indoor single person scenario and 0.964 in the outdoor single person scenario. Similarly, accuracies for indoor and outdoor multi-person scenarios rise to 0.925 and 0.896, respectively. The Combined scenario shows a notable improvement, reporting an accuracy of 0.93.

## 5 Conclusion

Throughout the course of this investigation, the authors faced several obstacles when attempting to address the various behaviors that were shown by real-time audiences while engaging with the system. Considering the difficulties that are linked with different environmental circumstances and different gestures, the system that has been suggested exhibits outstanding flexibility and strong durability. In this research work, effective design is developed that includes a user-friendly system able to handle members that come from a variety of different backgrounds by using cutting-edge models and methodologies. Gesture identification turned out to be an especially difficult issue, which led to the development of a wide variety of unique and diverse solutions. It was essential to use these tactics to reduce the possibility of faulty gesture recordings and to guarantee the dependability of the system. When it came to gender

recognition, the authors used a simple SVM that was trained on pre-existing data to make correct predictions about genders. In conclusion, the system that has been presented provides an extraordinary level of simplicity of use in real-time settings and maintains a constant delivery of highly accurate results throughout all phases, beginning with the detection of records and ending with their storage. It serves as a reliable and effective solution for managing healthcare scenarios, which are circumstances in which human involvement must be minimized to get the best possible results for public health. The resilience of the proposed system across different modules highlights its potential to transform the way in which companies engage with their consumers by enabling them to make choices based on data and adapt to changing requirements.

## References

1. Zhang F, Bazarevsky V, Vakunov A, Tkachenka A, Sung G, Chang CL, Grundmann M (2020) MediaPipe hands: on-device real-time hand tracking. arXiv.Org. <https://arxiv.org/abs/2006.10214>
2. Oudah M, Al-Naji A, Chahl J (2020) Hand gesture recognition based on computer vision: a review of techniques. J Imaging 6(8). <https://doi.org/10.3390/jimaging6080073>
3. Köpüklü O, Gunduz A, Kose N, Rigoll G (2019) Real-time hand gesture detection and classification using convolutional neural networks. arXiv.Org. <https://arxiv.org/abs/1901.10323>
4. Schlüsener N, Bücker M (2022) Fast learning of dynamic hand gesture recognition with few-shot learning models. arXiv.Org. <https://arxiv.org/abs/2212.08363>
5. Mahmud H, Morshed MM, Hasan Md K (2021) A deep learning-based multimodal depth-aware dynamic hand gesture recognition system. arXiv.Org. <https://arxiv.org/abs/2107.02543>
6. An X, Deng J, Guo J, Feng Z, Zhu X, Yang J, Liu T (2022) Killing two birds with one stone: efficient and robust training of face recognition CNNs by partial FC. arXiv.Org. <https://arxiv.org/abs/2203.15565>
7. Chen WL, Kang G, Huang PY, Chang X, Qian Y, Liang J, Gui L, Wen J, Chen P (n.d.) Argus: efficient activity detection system for extended video analysis
8. Anwar A, Raychowdhury A (2020) Masked face recognition for secure authentication. arXiv.Org. <https://arxiv.org/abs/2008.11104>
9. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) SSD: single shot multibox detector. In: Computer vision – ECCV 2016. Springer International Publishing, pp 21–37. [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2)
10. Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39(6):1137–1149. <https://doi.org/10.1109/tpami.2016.2577031>
11. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: 2016 IEEE Conference on computer vision and pattern recognition (CVPR). <https://doi.org/10.1109/cvpr.2016.91>
12. Wang CY, Bochkovskiy A, Liao HYM (2023) YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: 2023 IEEE/CVF Conference on computer vision and pattern recognition (CVPR). <https://doi.org/10.1109/cvpr52729.2023.00721>
13. Tan X, Triggs B (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: Analysis and modeling of faces and gestures. Springer Berlin Heidelberg, pp 168–182. [https://doi.org/10.1007/978-3-540-75690-3\\_13](https://doi.org/10.1007/978-3-540-75690-3_13)
14. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press

15. Wojke N, Bewley A, Paulus D (2017) Simple online and realtime tracking with a deep association metric. In: 2017 IEEE International conference on image processing (ICIP). <https://doi.org/10.1109/icip.2017.8296962>
16. Wojke N, Bewley A (2018) Deep cosine metric learning for person re-identification. In: 2018 IEEE Winter conference on applications of computer vision (WACV). <https://doi.org/10.1109/wacv.2018.00087>
17. Aharon N, Orfaig R, Bobrovsky BZ (2022) BoT-SORT: robust associations multi-pedestrian tracking.roy arXiv.Org. <https://arxiv.org/abs/2206.14651>
18. Agrawal A, Raj R, Porwal S (2013) Vision-based multimodal human-computer interaction using hand and head gestures. In: 2013 IEEE Conference on information and communication technologies. <https://doi.org/10.1109/cict.2013.6558300>
19. Sahoo JP, Ari S, Patra SK (2021) A user independent hand gesture recognition system using deep CNN feature fusion and machine learning technique. In: New paradigms in computational modeling and its applications. Elsevier, pp 189–207. <https://doi.org/10.1016/b978-0-12-822133-4.00011-6>
20. Dang TL, Tran SD, Nguyen TH, Kim S, Monet N (2022) An improved hand gesture recognition system using keypoints and hand bounding boxes. Array 16:100251. <https://doi.org/10.1016/j.array.2022.100251>

# Machine Learning Model for Cervical Cancer Risk Assessment



Vedavati Patil, Virendra Kumar Shrivastava, and Ashvini Alashetty

**Abstract** Cervical cancer is the leading cause of cancer related death among women worldwide and hence screening tools as well as risk prediction markers are very much needed. This study explored the prediction capacity of machine learning algorithms for cervical cancer risk factors with a large dataset. Description provided by the dataset: The data show information on diagnostic features, demographic data, and a clinical history of patients undergoing cervical cancer screening. There are two basic steps involved in data preprocessing, i.e., handling missing values, and standardizing numerical features pair plots and statistical summaries are the two approaches for exploratory data analysis (EDA) which can be used to understand how variables are distributed individually, or their relevance. The four machine learning models—XGBoost, Decision Tree, Random Forest, and Logistic Regression—are trained and assessed using GridSearchCV’s hyperparameter optimization. Performance metrics are used to evaluate how well the model predicts the risk of cervical cancer. These metrics include accuracy, precision, recall, and F1-score. Among the best are the Random Forest and XGBoost models, which demonstrate strong classification abilities and emphasize important variables like age, metrics related to sexual behavior, and cytological results. Random Forest achieved 94% accuracy, 60% precision, 54% recall, and 57% F1-score in contrast to XGBoost’s 94% accuracy, 53% precision, 63% recall, and 58% F1-score statistics. By integrating these models into an ensemble of voting classifiers, the predictive accuracy and reliability are improved for a variety of patient profiles.

**Keywords** Exploratory learning algorithms · Hyperparameter optimization · XGBoost · Random Forest · Logistic Regression · Decision Tree

---

V. Patil · V. K. Shrivastava (✉) · A. Alashetty

COE in Computer Vision, Department of Computer Science and Engineering, Alliance School of Advanced Computing, Alliance University, Bangalore, India

e-mail: [virendra.shrivastava@alliance.edu.in](mailto:virendra.shrivastava@alliance.edu.in)

V. Patil  
e-mail: [pvedavatibtech21@ced.alliance.edu.in](mailto:pvedavatibtech21@ced.alliance.edu.in)

A. Alashetty  
e-mail: [Ashvini.jagannath@alliance.edu.in](mailto:Ashvini.jagannath@alliance.edu.in)

## 1 Introduction

Cervical cancer is one of the severe health concerns, especially in settings with limited resources where access to routine screening is limited [10, 11]. Early detection is essential to improving treatment outcomes and lowering the death rates associated with this preventable disease [2]. Conventional screening techniques, like Pap smears, have proven successful but need infrastructure and expert interpretation that may not be available to everyone [19]. Through automated analysis of medical imaging data, recent advances in machine learning present promising avenues for improving the detection of cervical cancer [8, 9]. Machine learning models have the potential to enhance cervical cancer screening accessibility, efficiency, and accuracy by utilizing large datasets and complex algorithms [14].

### 1.1 Motivation and Problem Statement

Though progress has been made, there are still issues with correctly recognizing cervical abnormalities from medical images [20]. Interpretation of humans and may overlook early stage lesions that are essential for successful intervention. By offering unbiased, scalable, and possibly more precise diagnostic tools, machine learning techniques seek to overcome these drawbacks. This study explores the usage of machine learning (ML) algorithms for identification of cervical cancer from digitally captured medical images, motivated by the need for improved screening methods. Through an analysis of various models and methodologies, this study seeks to identify the most effective means of identifying risks in order to find trustworthy solutions that can enhance or supplement existing screening protocols.

### 1.2 Objectives of the Research

- Investigate the efficacy of ML models, specifically Random Forest (RF) and XGBoost, in detecting cervical abnormalities from image data.
- Assess the performance metrics such as model accuracy, recall, precision, and F1-score to quantify the efficacy of each model.
- Explore feature importance to identify key visual indicators and biomarkers associated with cervical cancer detection.
- Develop an ensemble model combining the strengths of individual classifiers to improve overall diagnostic accuracy.

## 2 Literature Review

Due to its high cost-effectiveness, the Pap smear test was chosen in this study over colposcopy [22]. The two crucial steps for classification in their work are the acquisition of a cell image and precise segmentation [1]. They will use ML models in this because manual detection takes a lot of time. Noise reduction has been achieved by bilateral filtering. To segment cells and nuclei, local Gaussian filtering energy segmentation is utilized; dice coefficient is used to measure accuracy; contour marking is accomplished by adjusting parameters like epsilon, alpha, lamda1, lamda2, and iteration count. PCA was utilized to lower the feature set's dimensionality. Lastly, three different multi-class SVM classifier types (line, polynomial, and quadratic for image classification, Gaussian RBF) were employed. By using dataset they used, they have demonstrated that SVM is the most effective classifier. Twenty photos were used to test the model and 200 images were used to train the classifier. For each classifier, confusion matrices were plotted to assess the performance of the model. Polynomial SVM yielded the highest accuracy of 95%.

The growth of cervical cancer ranks 4<sup>th</sup> among the most common diseases in females. Thus, they have suggested machine learning models like Decision Trees in this paper. To predict cervical cancer and its variables, Random Forest and XGBooster are used [2, 24]. By comparing other methods, they have increased their accuracy. The performances of these classifiers were assessed by using matrices (F1-score, precision, accuracy, and recall). They have classified data into cervical cancerous and no cervical cancerous where they got 93.33% accuracy for Decision Tree with 93.33% accuracy for Random Forest, and 93.33 for XGBooster. These classifiers not only work well with accuracy but also with the other three matrices. The future work will be focused on improving the performance of these classifiers in detecting cervical cancer by adding advanced techniques like deep learning [25–29].

Pap smear slides are used in this paper's application of the Mask Regional Convolution Neural Network for the screening of cervical cancer [3]. By identifying and examining the nucleus of the cervical cells—where the slides include both cervical cells and different artifacts like white blood cells—they hope to differentiate between the normal and abnormal nuclear classes. The suggested algorithm produced results with 91.7% accuracy, 57.8% precision, 91.7% sensitivity, and 57.8% specificity. Future research may combine the mask RCNN and modified Deep Pap model, which would use slides with images of complete Pap smears as input to locate the nucleus and to discover cell types. The modified Deep Pap algorithm used the position of nucleus as an input. They have investigated the effect on Pap smear slide images using mask RCNN model to classify normal and abnormal cells.

Since cervical cancer can spread to other part of body, early detection is crucial. The Pap smear test is recommended over the colposcopy test because it is less expensive and painless. In order to create a benchmark for evaluating upcoming classification techniques [21], the research applies deep learning classification techniques to the SIPAKMED Pap smear image data set [3]. With the Rest-152 architecture, accuracy of 94.89% was attained. Five classes—Dykerotolic, Metaplastic, Koilocytotic,

Parapusal, and superficial intermediate—were used for the classification process. The hardest performance among these classifiers was achieved by Rest-net 152. Future work on this project may involve incorporating multiple datasets with similar Pap smear modalities to extrapolate the results more robustly.

Using convolution neural networks (CNNs), the classification of cervical cancer cells was accomplished in this paper [4]. To extract features for image classification, the cell images were fed into convolution neural network models; an extreme learning machine has been utilized. They suggested an extremely accurate CNN-ELM system that achieves 91.2% accuracy in seven classes and 99.5% accuracy in two classes. The Bethesda system introduces cancer diagnosis using inter observer variability [5] which was time-consuming. To make efficient predictions, they developed a deep learning screening system. The applied six distinct deep CNNs—VGGNet, AlexNet, VGG-16 and VGG-19, ResNet-50 and ResNet-101, and GoogleNet—are examined in this study on computer-aided screening [13]. It emphasizes how adding three of the best models to an ensemble classification can produce multi-class classification with high accuracy. Three distinct data sets—liquid-based cytology, conventional, and Herlex data sets—are used to assess their suggested methodology. They achieved 0.99% precision with the ensemble classifier. Allehaibi et al. intend to create automatic techniques for the classification of cervical cancer by segmenting and classifying data using mask regional neural networks (mask RCNNs) and classifying data using VGG-like networks, like Net [6]. The best segmentation results, with  $0.92 \pm 0.06$  precision, and  $0.91 \pm 0.05$  recall, were found when using mask RNN. They also implemented two classification scenarios. William et al. described the PAT tool for cervical cancer automated classification from Pap smear images, as the conventional methods are time-consuming [7].

Utilizing deep learning models to categorize cervical cells in order to improve monitoring and cervical cancer early detection. It will be easier to integrate the good and bad aspects if you use the SSD model. With the addition of complementary features, the suggested SSD network increases overall accuracy and sensitivity, making it appropriate for cell classification-based early automatic detection of cervical cancer [8]. The paper addresses the inadequate sensitivity of classical networks to small objects and presents an enhanced SSD network. Identifies tiny cells in complicated backgrounds with effectiveness. This model has a mean average precision of 81.53% and an accuracy of 90.8%. Compared to interclass differences, this has a better ability to handle intra class differences. These features also maintain quick detection times while enhancing sensitivity and overall accuracy. Demonstrates a 4.92% increase in MAP and a 7.54% increase in accuracy over baseline and state-of-the-art models. Plan to train the network on bigger datasets in order to enhance accuracy and generalization performance, supporting automated cervical cancer detection.

Due to a lack of access to high-quality healthcare, cervical cancer is the fourth deadly disease in developing nations [9]. This study describes a fully automated deep learning pipeline that uses cervigram images to classify cervical cancer and detect cervixes. The pipeline makes use of two pre-trained models: One with an AUC score of 0.82 for tumor classification and another with an IoU of 0.68 for

cervix detection [23]. This system is perfect for mobile deployment in resource-constrained environments to improve early cancer detection, as it has the ability to detect cancer up to 1000 times faster and classify data up to 20 times faster than current models.

### 3 Methodology

#### 3.1 Data Preprocessing

- **Dataset:** The study employed a dataset that includes [UCI repository dataset], and important characteristics related to the detection of cervical cancer are discussed in the result section.
- **Processing Steps:** Initial missing values in the dataset were denoted by ‘?’; these were subsequently consistently replaced with NaN values. Then, for analysis, the numerical columns were transformed into the proper data types. To comprehend the feature distribution and spot possible outliers, simple statistical summaries were created. To investigate correlations and spot any patterns among numerical variables, visual inspections using pair plots were carried out.
- **Handling Missing Values, Scaling, and Encoding:** The strategies used to impute missing values were specific to each type of data. For example, numeric features were scaled using Standard Scaler to normalize their distributions and the median value was used to impute them. To effectively handle categorical data, one-hot encoding was used to encode categorical variables with the most frequently occurring values.

#### 3.2 Machine Learning Model

Based on how well-suited each classification model was for the job of detecting cervical cancer, four were chosen: Based on the suitability of classification models for detecting cervical cancer, four classification models were selected:

- **Logistic Regression (LR):** Baseline model known for its interpretability and simplicity is Logistic Regression [16].
- **Decision Trees (DT):** Complex interactions within data can be captured by non-linear models [15, 18].
- **Random Forest (RF):** Decision Tree-based ensemble techniques that provide strong performance through aggregation.
- **XGBoost:** Gradient boosting algorithms are well-known for their ability to handle big datasets quickly and increase the precision of predictions.

**Feature Engineering:** In order to improve model performance, feature engineering entailed extracting pertinent features from the dataset and modifying them. Although it isn't stated specifically, feature importance analysis was done after modeling to find important factors affecting the classification of cervical cancer.

**Model Training:** To guarantee uniformity in the way that data was treated across models, a pipeline approach that combined preprocessing stages (imputation, scaling, and encoding) was used to train each model. During the model's training process, cross-validation techniques were used to maximize performance metrics like accuracy, precision, recall, and F1-score. GridSearchCV was used to find the ideal parameters for the Random Forest and XGBoost models, improving their predictive power through hyperparameter tuning.

### 3.3 Ensemble Techniques

**Description of Ensemble Approach:** An ensemble technique for Voting Classifiers was used to improve predictive performance even further. By combining predictions from the top-performing models (XGBoost and Random Forest), this method reduced the biases in each model individually and increased classification accuracy overall [17].

**Justification for Ensemble Methods:** The ensemble method of Voting Classifiers was supported by empirical data from the individual model performances. The objective of the ensemble approach was to improve the reliability of cervical cancer detection by achieving higher classification accuracy and robustness against over fitting. This was achieved by utilizing the strengths of different algorithms, namely XGBoost for improved predictive power and Random Forest for robustness.

## 4 Result

The standard metrics for classification tasks: Accuracy, F1-score, precision, and recall, were used to assess the machine learning models.

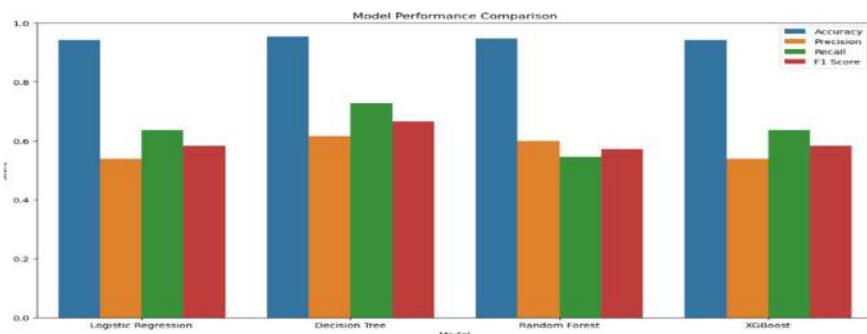
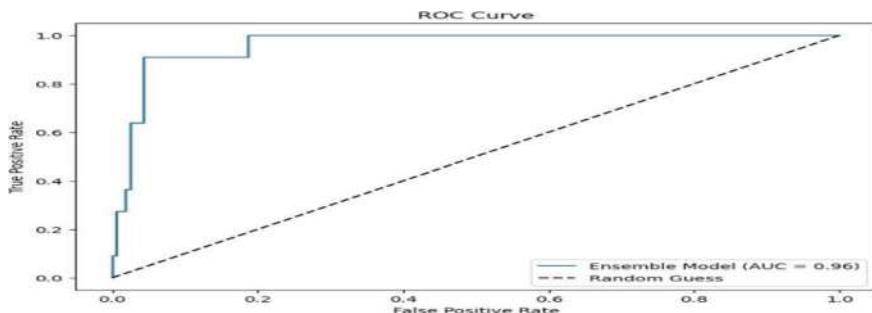
Table 1 presents performance metrics of the different model. Comparison graph of all algorithms is shown in Fig. 1. The graph gives details about all the algorithms by comparing with performance metrics.

### 4.1 ROC Curves and AUC Scores of Ensemble Methods

Figure 2 presents ROC curve of the ensemble model.

**Table 1** Performance metrics of the different model

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
LR	94	53	63	58
DT	95	61	72	66
RF	94	60	54	57
XGB	94	53	63	58

**Fig. 1** Comparison graph of all algorithms. The graph gives details about all the algorithms by comparing with accuracy, precision, recall, F1-score**Fig. 2** ROC curve. The graph gives details about the ensemble model AUC = 0.96

The contribution of each attribute was examined for the RF and XGBoost models. The following is a visualization of the key characteristics affecting the classification:

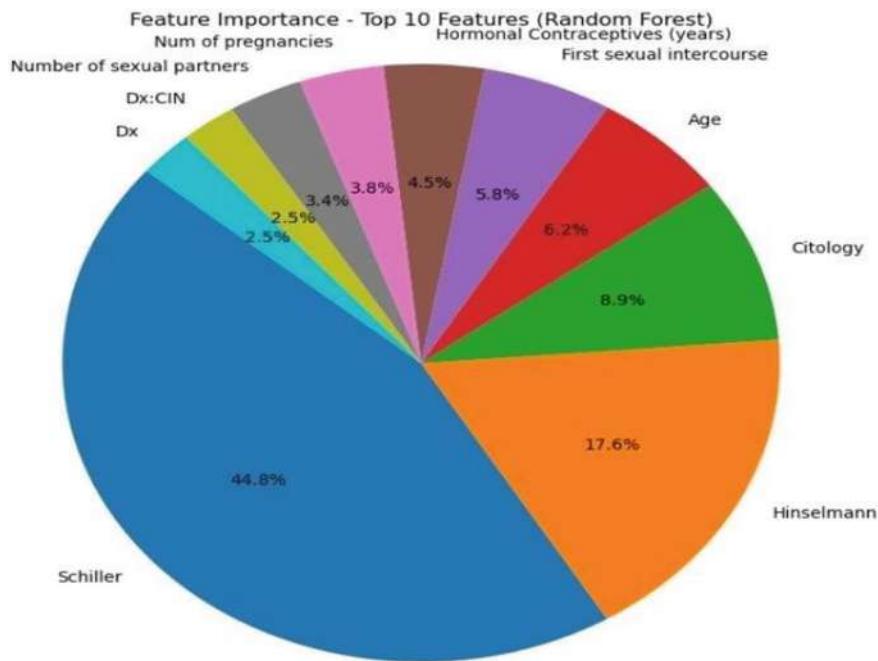
- RF
- XGBooster

## 5 Results and Discussion

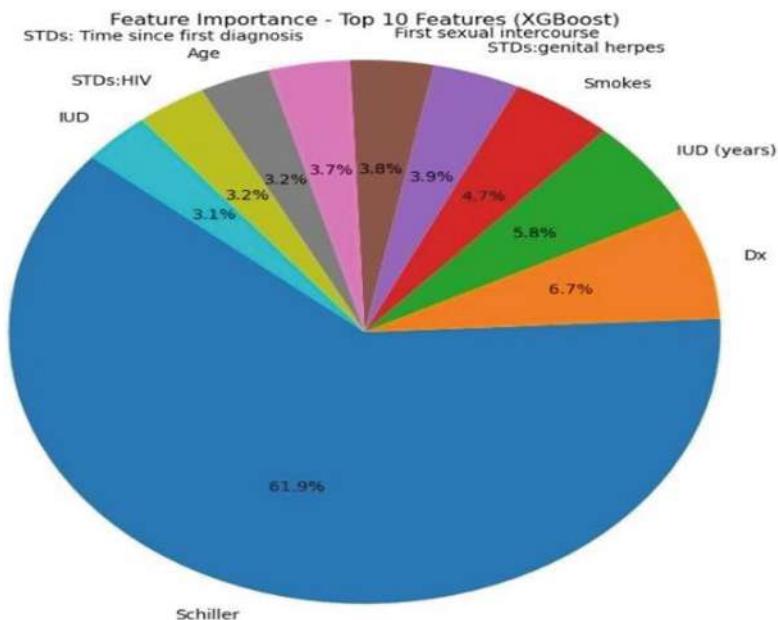
Out of all the models evaluated, the XGBoost model showed the best performance, as shown by its highest accuracy, precision, recall, and F1-score. This suggests that the accuracy of cervical cancer detection prediction is enhanced when gradient boosting and Random Forest techniques are combined. Benefits and drawbacks of the models Logistic Regression while simple to use, its performance metrics were lower than those of ensemble methods, suggesting that it is not as effective in handling non-linear relationships in data. Decision Tree: Showed competitive performance, but its generalizability was limited by its propensity to over fit on complicated datasets. Random Forest: Using ensemble averaging to increase model stability and performance metrics, it demonstrated resilience against over fitting. XGBoost: Outperformed other models because it could improve weak learners one after the other, increasing model robustness and predictive accuracy. The study emphasizes the value of ensemble methods and feature engineering in enhancing cervical cancer detection classification accuracy. The crucial role that particular biomarkers and clinical characteristics play in differentiating between cancerous and non- cancerous conditions is one of the major discoveries. Our results are consistent with previous research showing the effectiveness of gradient boosting algorithms and ensemble approaches in cervical cancer diagnostics [12]. Additionally, the study makes a contribution by verifying these methods on a particular dataset related to cervical cancer, thereby bolstering their suitability for use in clinical settings. The top 5 risk factors (features) that proposed models have identified as most important in predicting the likelihood of cervical cancer based on the feature importance analysis from your Random Forest and XGBoost models are presented in Figs. 3 and 4, respectively.

## 6 Conclusion

The application of ML models: LR, DT, RF, and XGBoost for cervical cancer detection was examined in this study. When compared to individual models, the results show that ensemble techniques—in particular, XGBoost—perform better in terms of accuracy. Proving the usefulness of ensemble approaches for the detection of cervical cancer in medical image analysis. Giving clinical decision-makers insights into the model performance metrics and feature importance. Highlighting the potential for ML models to enhance cervical cancer early detection and classification. In the future, our work can be done by adding a wider range of datasets to improve the model's generalizability across various demographic groups and geographic areas. Investigating deep learning architectures and sophisticated ensemble methods to increase classification accuracy even more.



**Fig. 3** Ranks of risk factors with Random Forest (The pie chart provides the top 5 risk factors which causes cervical cancer)



**Fig. 4** Ranks of risk factors with XGBooster (The pie chart provides the top 5 risk factors which causes cervical cancer)

## References

- Arora A, Tripathi A, Bhan A (2021) Classification of cervical cancer detection using machine learning algorithms. In: 2021 6th International conference on inventive computation technologies (ICICT), pp 827–835. IEEE
- Akter L, Islam MM, Al-Rakhami MS, Haque MR (2021) Prediction of cervical cancer from behavior risk using machine learning techniques. *SN Comp Sci* 2(3):177
- Sompawong N, Mopan J, Pooprasert P, Himakhun W, Suwannaruk K, Ngamvirojcharoen J, Tantibundhit C (2019) Automated pap smear cervical cancer screening using deep learning. In: 2019 41st Annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 7044–7048. IEEE
- Ghoneim A, Muhammad G, Hossain MS (2020) Cervical cancer classification using convolutional neural networks and extreme learning machines. *Futur Gener Comput Syst* 102:643–649
- Hussain E, Mahanta LB, Das CR, Talukdar RK (2020) A comprehensive study on the multi-class cervical cancer diagnostic prediction on pap smear images using a fusion-based decision from ensemble deep convolutional neural network. *Tissue Cell* 65:101347
- Allehaibi KHS, Nugroho LE, Lazuardi L, Prabuwono AS, Mantoro T (2019) Segmentation and classification of cervical cells using deep learning. *IEEE Access* 7:116925–116941
- William W, Ware A, Basaza-Ejiri AH, Obungoloch J (2019) A pap-smear analysis tool (PAT) for detection of cervical cancer from pap-smear images. *Biomed Eng Online* 18:1–22
- Jia D, Zhou J, Zhang C (2022) Detection of cervical cells based on improved SSD network. *Multimedia Tools Appl* 81(10):13371–13387
- Alyafeai Z, Ghouti L (2020) A fully-automated deep learning pipeline for cervical cancer classification. *Expert Syst Appl* 141:112951

10. Ndikom CM, Ofi BA (2012) Awareness, perception and factors affecting utilization of cervical cancer screening services among women in Ibadan, Nigeria: a qualitative study. *Reprod Health* 9:1–8
11. Lu J, Song E, Ghoneim A, Alrashoud M (2020) Machine learning for assisting cervical cancer diagnosis: an ensemble approach. *Futur Gener Comput Syst* 106:199–205
12. Bahad P, Saxena P (2020) Study of adaboost and gradient boosting algorithms for predictive analytics. In: International conference on intelligent computing and smart communication 2019: Proceedings of ICSC 2019, pp 235–244. Springer Singapore
13. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 25
14. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, Bray F (2019) Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer* 144(8):1941–1953
15. Vos D, Verwer S (2021) Efficient training of robust decision trees against adversarial examples. In: International conference on machine learning, pp 10586–10595. PMLR
16. Long WJ, Griffith JL, Selker HP, D'agostino RB (1993) A comparison of logistic regression to decision-tree induction in a medical domain. *Comput Biomed Res* 26(1):74–97
17. Ali MS, Hossain MM, Kona MA, Nowrin KR, Islam MK (2024) An ensemble classification approach for cervical cancer prediction using behavioral risk factors. *Healthc Anal* 5:100324
18. Nagy K, Reiczigel J, Harnos A, Schrott A, Kabai P (2010) Tree-based methods as an alternative to logistic regression in revealing risk factors of crib-biting in horses. *J Equine Vet Sci* 30(1):21–26
19. Simonyan K (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
20. Bora K, Chowdhury M, Mahanta LB, Kundu MK, Das AK (2017) Automated classification of Pap smear images to detect cervical dysplasia. *Comput Methods Programs Biomed* 138:31–47
21. Norup J (2005) Classification of Pap-smear data by tranduction neuro-fuzzy methods (Master's thesis, Technical University of Denmark, DTU, DK-2800 Kgs. Lyngby, Denmark)
22. Yoo TK, Choi JY, Kim HK (2020) A generative adversarial network approach to predicting postoperative appearance after orbital decompression surgery for thyroid eye disease. *Comput Biol Med* 118:103628
23. Adem K, Kılıçarslan S, Cömert O (2019) Classification and diagnosis of cervical cancer with stacked autoencoder and softmax classification. *Expert Syst Appl* 115:557–564
24. Wankhede DS, Shelke CJ, Shrivastava VK, Achary R, Mohanty SN (2024) Brain tumor detection and classification using adjusted InceptionV3, AlexNet, VGG16, VGG19 with ResNet50–152 CNN Model. EAI Endorsed Trans Pervasive Health Technol 10
25. Shrivastava VK, Shelke CJ, Shrivastava A, Mohanty SN, Sharma N (2023) Optimized deep learning model for disease prediction in potato leaves. EAI Endorsed Trans Pervasive Health Technol 9
26. Shrivastava VK, Shrivastava A, Sharma N, Mohanty SN, Pattanaik CR (2023) Deep learning model for temperature prediction: a case study in New Delhi. *J Forecast* 42(6):1445–1460
27. Saini V, Rai N, Sharma N, Shrivastava VK (2022) A convolutional neural network based prediction model for classification of skin cancer images. In: International conference on intelligent systems and machine learning, pp 92–102. Cham: Springer Nature Switzerland
28. Singhal A, Phogat M, Kumar D, Kumar A, Dahiya M, Shrivastava VK (2022) Study of deep learning techniques for medical image analysis: a review. *Mater Today: Proc* 56:209–214
29. Shrivastava VK, Kumar A, Shrivastava A, Tiwari A, Thiru K, Batra R (2021) Study and trend prediction of Covid-19 cases in India using deep learning techniques. *J Phys: Conf Ser* 1950(1):012084. IOP Publishing

# **H<sub>∞</sub>-Based Fractional-Order Controller for Trajectory Tracking Control of Nonholonomic Mobile Robots**



**Km Shelly Chaudhary and Naveen Kumar**

**Abstract** This paper examines the nonholonomic mobile robot's trajectory tracking control problem, considering external disturbances and the dynamical system's uncertainties. The offered control technique presents a sliding surface of fractional-order to provide a quicker response from the controller. Utilizing the presenting sliding surface, a robust  $H_{\infty}$ -based fractional-order sliding mode controller is presented to efficiently handle these external influences to maintain the robust performance of the dynamical structure. The controller design employs a model-based control approach in conjunction with a  $H_{\infty}$  tracking control technique to reduce the negative effects of the uncertainties and external disturbances. Lyapunov's stability criteria is used to demonstrate the stability of the dynamical system. Additionally, the numerical simulation results demonstrate the efficacy of the designed controller in a comparative manner.

**Keywords** Nonholonomic mobile robots · Fractional-order sliding surface ·  $H_{\infty}$  tracking control · Lyapunov approach

## **1 Introduction**

Owing to its multiple roles in the medical field, organizations, military activities, and many other industries, position-tracking control of nonholonomically constrained mobile robot has garnered significant interest recently [1–3]. So by seeing their application-specific utility, the main motivation of this work is to develop an efficient control technique to handle mobile robots more appropriately in a robust manner. The

---

K. S. Chaudhary (✉)

Meerut College, Meerut, Uttar Pradesh, India

e-mail: [shellymath23@gmail.com](mailto:shellymath23@gmail.com)

N. Kumar

Mahatma Jyotiba Phule Rohilkhand University Bareilly, Bareilly, Uttar Pradesh, India

K. S. Chaudhary · N. Kumar

National Institute of Technology, Kurukshetra, Haryana, India

dynamical structure of nonholonomic mobile robots is extremely nonlinear, coupled, and time-varying. So, managing these nonlinear properties, uncertainties, outside disturbances, etc., presents a number of practical difficulties while controlling them. Numerous classical controllers, including PID controllers, back-stepping controllers, adaptive controllers, sliding surface-based controllers (SMCs), and many others [4–6], have been proposed in the literature.

To effectively execute the dynamical system, many combinations of sliding mode control techniques with fractional-order controllers have been described in the literary texts as ways to improve and accurately perform the intended tasks. As compared to integer-order controllers, fractional-order controllers perform better due to their faster order convergence speed [7]. Applying the fractional-order derivative to linear sliding mode controllers, also referred to as the fractional-order SMC, is the first step in the work on integrating the fractional-order derivative with sliding mode controllers [8, 9]. Compared to traditional sliding mode controllers, these controllers have superior tracking capability. Using a time-dependent sliding manifold, Eray et al. [10] present a fractional-order SMC technique. While fractional-order sliding mode controllers aid in the dynamic system's effective operation, the system's overall performance deteriorates because of the detrimental effects of external factors and system uncertainties. Panwar [11] presents an  $H_\infty$ -based trajectory tracking control approach for robot manipulators utilizing the terminal SMC technique. The integration of  $H_\infty$ -based controller with sliding mode controller gave an outstanding performance of dynamical system with controlled gain values and significantly faster response of system states toward their trajectories.

In this queue, this study presents a novel combination of fractional-order controllers, sliding mode-based controllers, model-based controllers, and  $H_\infty$ -based controllers for enhanced performance of nonholonomic mobile robots, inspired by the previous work. The following are the work's primary contributions.

- (1) In the context of outside disruptions and system uncertainties, the position-tracking problem of a nonholonomic wheeled mobile robot system is investigated.
- (2) An elegant combination of fractional-order sliding mode control approach with model-based robust  $H_\infty$  trajectory tracking control approach is proposed based on a fractional-order sliding manifold.
- (3) The convergence of tracking errors toward equilibrium and the dynamical system's stability are examined by using the Lyapunov stability criterion.
- (4) Using a 3-dof nonholonomic robot manipulator system, a simulation study is carried out for the suggested approach, and the outcomes are compared with the existing techniques.

The remaining part of this work is given: The dynamical model of the system is given in Sect. 2. The control strategy is explained in Sect. 3, and the stability analysis is analyzed in Sect. 4. The simulation results are presented in Sect. 5, and the work is concluded in Sect. 6.

## 2 Dynamical Model of Nonholonomic Wheeled Mobile Robot

For the nonholonomic mobile robotic system, the Euler-Lagrange dynamical equation is expressed as follows:

$$M(\theta)\ddot{\theta} + V_Q(\theta, \dot{\theta})\dot{\theta} + F(\dot{\theta}) + \tau_d(t) = B(\theta)U + Q^T(\theta)\lambda_Q \quad (1)$$

where  $\theta = [x, y, \Theta]^T \in R^{3 \times 1}$  denotes the coordinates for mobile base,  $M(\theta) \in R^{3 \times 3}$  denotes the inertial matrix,  $V_Q(\theta, \dot{\theta}) \in R^{3 \times 3}$  gives the centripetal or coriolis matrix,  $F(\dot{\theta}) \in R^{3 \times 1}$  denotes friction to the dynamical system,  $\tau_d(t) \in R^{3 \times 1}$  denotes the bounded disturbance,  $B(\theta) \in R^{3 \times 2}$  transforms the states of mobile manipulator system in the earth's reference frame,  $U \in R^{3 \times 1}$  denotes the control torque input,  $Q^T(\theta) \in R^{3 \times 1}$  denotes the constraint matrix associated with  $\lambda_Q \in R$  as Langranges multiplier.

Assume that the following nonholonomic kinematic restriction applies to the mobile robot system as

$$Q(\theta)\dot{\theta} = 0 \quad (2)$$

Equation (1) of mobile robots to the manifold  $\mathfrak{J}_Q$  is limited by these restrictions, which may be represented as:  $\mathfrak{J}_Q = \{(\theta, \dot{\theta}) | Q(\theta)\dot{\theta} = 0\}$ . Equation (2) gives us the complete rank matrix  $T(\theta) \in R^{3 \times 2}$ , which results in this:

$$T^T(\theta)Q^T(\theta) = 0 \quad (3)$$

A new vector  $\chi = [y, \Theta]^T \in R^2$  that meets the following condition is obtained from the restrictions stated in Equations (2) and (3).

$$\dot{\theta} = T(\theta)\dot{\chi} \quad (4)$$

Diff. it, we may get

$$\ddot{\theta} = T(\theta)\ddot{\chi} + \dot{T}(\theta)\dot{\chi} \quad (5)$$

Integrating Equations (4) and (5) in Equation (1), and then multiplying the resulting equation by  $T^T$ .

$$\bar{M}_g\ddot{\chi} + \bar{V}_g\dot{\chi} + \bar{F}_g + \bar{\tau}_{gd} = T^T U \quad (6)$$

with  $\bar{M}_g = T^T M(\theta)T$ ,  $\bar{V}_g = T^T M(\theta)\dot{T} + T^T V_Q(\theta, \dot{\theta})T$ ,  $\bar{F}_g = T^T F(\dot{\theta})$ ,  $\bar{\tau}_{gd} = T^T \tau_d$ .

Let the following properties and assumptions be satisfied by the aforesaid constrained dynamics equation (6) of nonholonomic mobile robots.

**Proposition 1** The matrix  $\bar{M}_g$  is symmetric, bounded, and has all its eigenvalues greater than zero.

**Proposition 2** The dynamical terms  $\bar{F}_g \leq g_1 + g_2 \|\dot{\chi}\|$  and  $\|\bar{\tau}_{gd}\| \leq g_3$  is bounded for arbitrary constants  $g_i > 0$  for  $i = 1, 2, 3$ .

**Assumption 1** Every Jacobian matrix is continuous and uniformly bounded if  $\chi = [y, \Theta]^T \in R^2$  is continuous and uniformly bounded.

### 3 Controller Design

#### 3.1 Surface

The sliding surface be chosen as

$$r(t) = D^{\alpha+1} \bar{\chi}(t) + \dot{\bar{\chi}}(t) + A \bar{\chi}(t) \quad (7)$$

Here  $0 < \alpha < 0$ ,  $\bar{\chi} = \chi_d - \chi$  gives the position/location tracking errors,  $\chi_d(t) \in R^2$  gives the reference path,  $A = Diag[A_1, A_2] \in R^{2 \times 2}$  having  $A_1, A_2$  as positive values, and  $r(t) = [r_1(t), r_2(t)]^T \in R^2$  be sliding variable. The  $k^{th}$  point of the above surface given in Eq. (7) be expressed as

$$r_k(t) = D^{\alpha+1} \bar{\chi}_k(t) + \dot{\bar{\chi}}_k(t) + A \bar{\chi}_k(t) \quad (8)$$

with  $k = 1, 2$ .

Differentiating Eq. (8) and rewriting it in the explicit form, we get

$$\dot{r}(t) = D^{\alpha+2} \bar{\chi}(t) + \ddot{\chi}_d(t) - \ddot{\chi}(t) + A \dot{\bar{\chi}}(t) \quad (9)$$

#### 3.2 $H_\infty$ Robust Control

To overcome the negative influence of the system uncertainties and outside disturbances in a robust manner, the following  $H_\infty$ -based robust tracking approach is presented. For the appropriate choice of  $\psi > 0$  and  $T \in (0, \infty)$ , the attenuate behavior of the control approach can be gained if the dynamical structure of nonholonomic wheeled mobile robot satisfies the following condition:

$$\int_0^T \bar{\chi}^T W_1 \bar{\chi} \leq r(0)^T \bar{M}_{g0} r(0) + \bar{\chi}(0)^T W_2 \bar{\chi}(0) + \psi^2 \int_0^T \|\bar{\tau}_{gd}\|^2 dt \quad (10)$$

where  $W_1$  and  $W_2$  are symmetric positive definite matrices.

### 3.3 Proposed Controller

Let the proposed control input torque be given as

$$T^T U = U_{\text{smc}} + U_1 \quad (11)$$

where  $U_{\text{smc}}$  and  $U_1$  are given as follows:

$$U_{\text{smc}} = \bar{M}_g(D^{\alpha+2}\bar{\chi}(t) + \ddot{\chi}_d(t) + A\dot{\bar{\chi}}(t)) + \bar{V}_g\dot{\bar{\chi}} + \bar{F}_g + \bar{\tau}_{gd} \quad (12)$$

and

$$U_1 = \bar{M}_g(W_3r(t) + W_4\text{sign}(r(t)) + Q\bar{\chi} + \frac{r(t)}{2\psi^2}) \quad (13)$$

with  $W_3$  and  $W_4$  be symmetric positive definite matrices.

## 4 Stability Analysis

**Theorem 1** *The trajectory tracking errors converge to their equilibrium states, and the signals' boundedness is achieved if we choose the condition given in Eq. (10) and the control input torque as given in Eq. (11).*

**Proof** Considering the Lyapunov function candidate as

$$L = \frac{1}{2}r(t)^T r(t) + \frac{1}{2}\bar{\chi}(t)^T W_2\bar{\chi}(t) \quad (14)$$

Differentiate Equation (14) with respect to  $t$  and using Equation (9), we get

$$\begin{aligned} \dot{L} = & r(t)^T (D^{\alpha+2}\bar{\chi}(t) + \ddot{\chi}_d(t) + A\dot{\bar{\chi}}(t) + \bar{M}_g(-T^T U + \bar{V}_g\dot{\bar{\chi}} \\ & + \bar{F}_g + \bar{\tau}_{gd})) + \bar{\chi}(t)^T W_2\dot{\bar{\chi}}(t) \end{aligned} \quad (15)$$

Putting the value of  $\int_0^T \bar{\chi}^T W_1 \bar{\chi}$  and  $T^T U$  from Equations (10) and (11). To simplify, we will have

$$L(T) - L(0) \leq \frac{1}{2}\psi^2 \int_0^T \|\bar{\tau}_{gd}\|^2 - \int_0^T \bar{\chi}^T W_5 \bar{\chi} \quad (16)$$

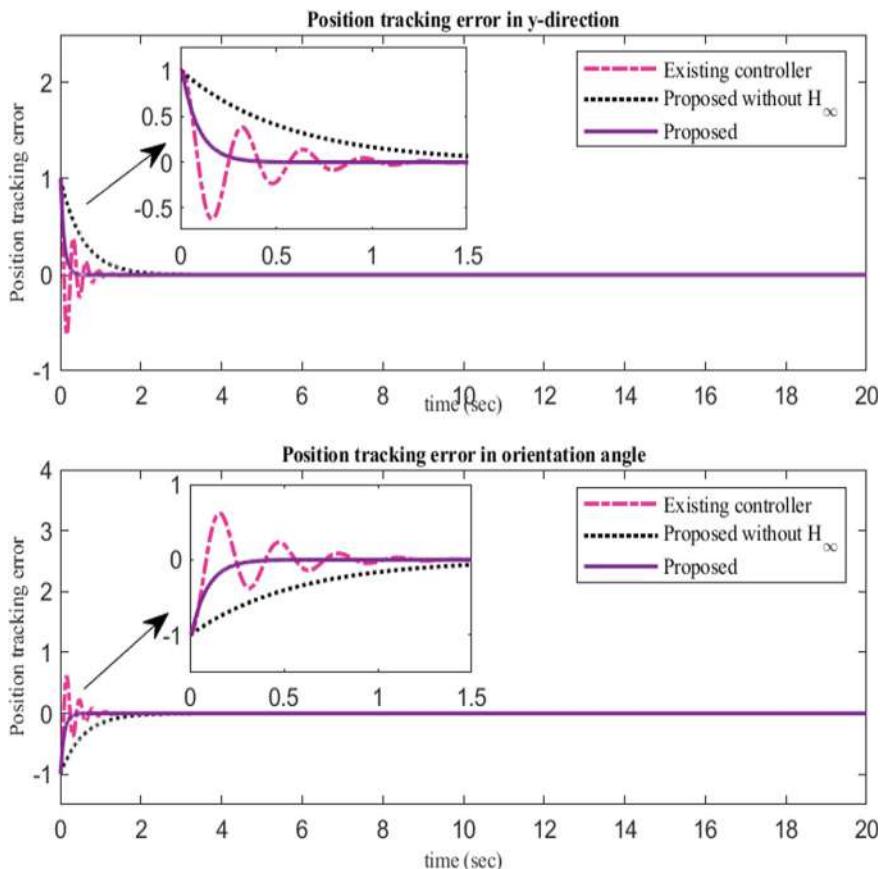
As  $\|\bar{\tau}_{gd}\| \in L^2[0, t]$  for  $t \geq 0$  implies  $L(\bar{\chi}, r(t))$  is bounded function. Hence, it can be concluded that  $r(t)$  and  $\bar{\chi}$  are bounded. By Lyapunov stability criteria, it has been concluded that the tracking errors go to their equilibrium states with the stability of the overall dynamical system.

## 5 Simulation

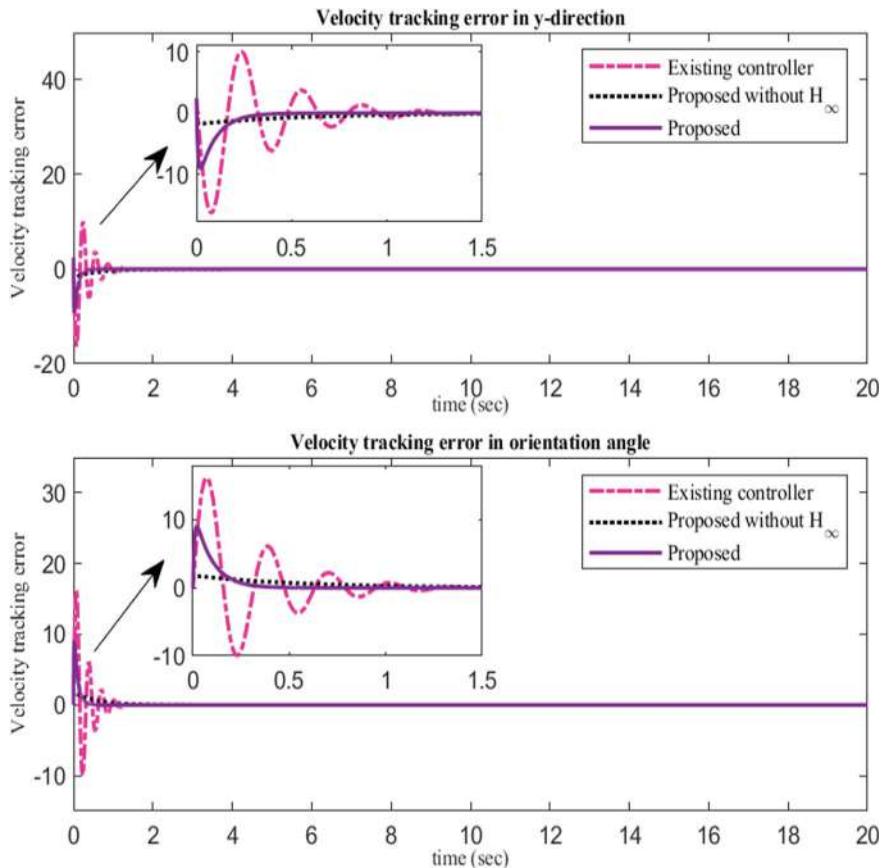
The dynamical model of a 3-dof nonholonomic mobile robot [12] is used in the simulation study to show the proposed controller's effective performance. The nonholonomic restriction is expressed as follows for a mobile robot system:  $-\dot{x} \cos(\theta) + \dot{y} \sin(\theta) = 0$ . The Matlab ODE45 solver is used to conduct the simulation investigation on the nonholonomic mobile robot. Grunwald-Letnikov(GL) derivative definition [11] has been used to calculate fractional-order derivatives.

The reference trajectories are chosen as follows:  $y_d = 1.5 \sin(t) + \cos(t) + t$  and  $\Theta_d = 0.25\pi(1 - \cos(t)) - 1$  with  $[y_d(0), \Theta_d(0)] = [0, 0]$  and external disturbances as  $\bar{\tau}_{gd} = [0.1 \sin(t); 0.1 \cos(t)]$ .

Simulation results are conducted for three cases: existing controller, proposed controller without  $H_\infty$ , and proposed controller, to show the superiority of the designed

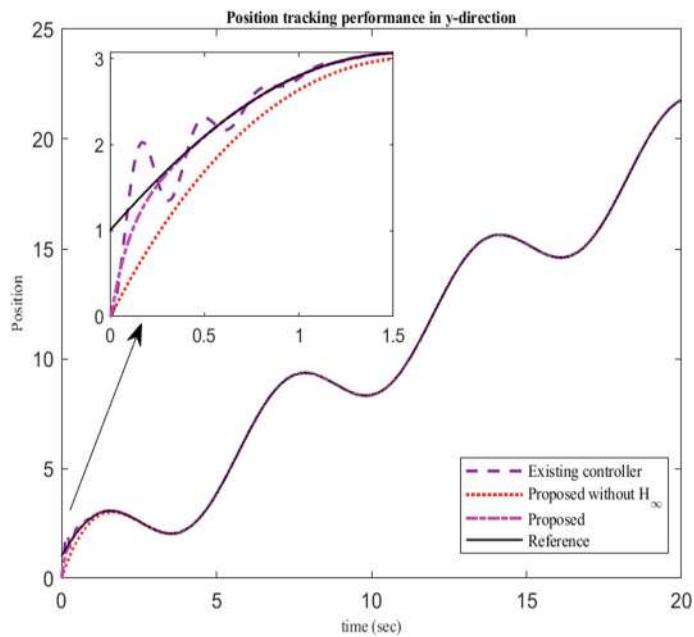


**Fig. 1** Position-tracking errors

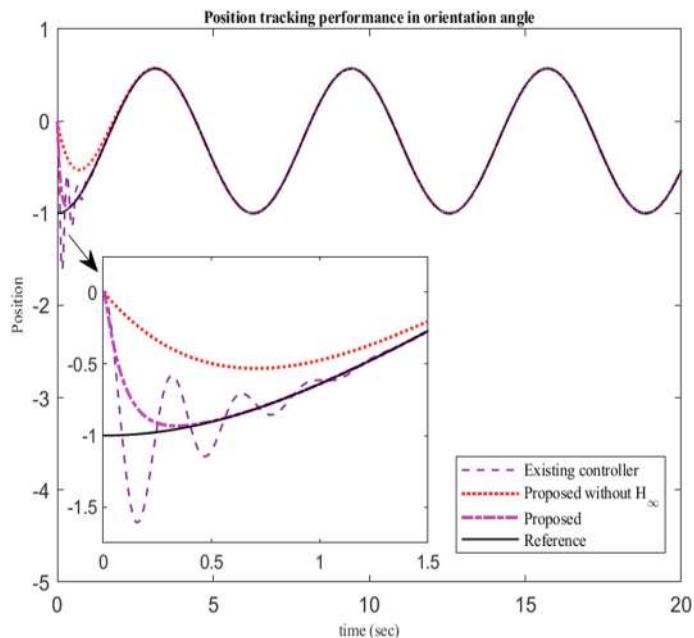


**Fig. 2** Velocity tracking errors

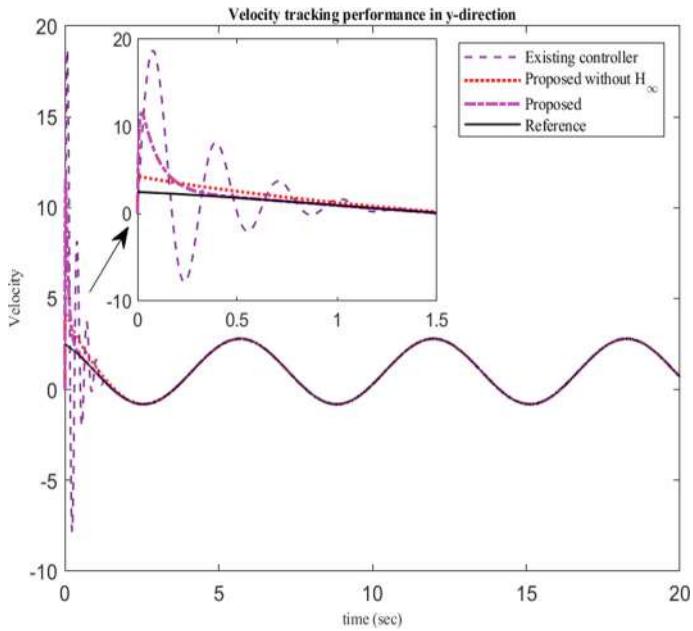
control technique. The simulated results in the first case are obtained using the method described in the article [10] while in the second case, the results are performed with the proposed control approach without  $H_{\infty}$  control to illustrate the robustness of the offered controller. In the last case, the behavior of the dynamical system with the proposed control approach is presented. The initial conditions and parameters are taken the same for all these three cases. The performance of the offered controller for a nonholonomic mobile robot system is given in Figs. 1, 2, 3, 4, 5, and 6 in a comparative manner. Figures 1 and 2 show both generalized coordinates's position and velocity tracking errors demonstrating the quicker convergence of tracking errors for the proposed control scheme within a very small settling time to reach the reference trajectory as compared to the other cases. In Figs. 3 and 4, the position-tracking performance is demonstrated in a comparative manner. The efficiency with which the dynamic system tracks the reference trajectories is demonstrated by these figures



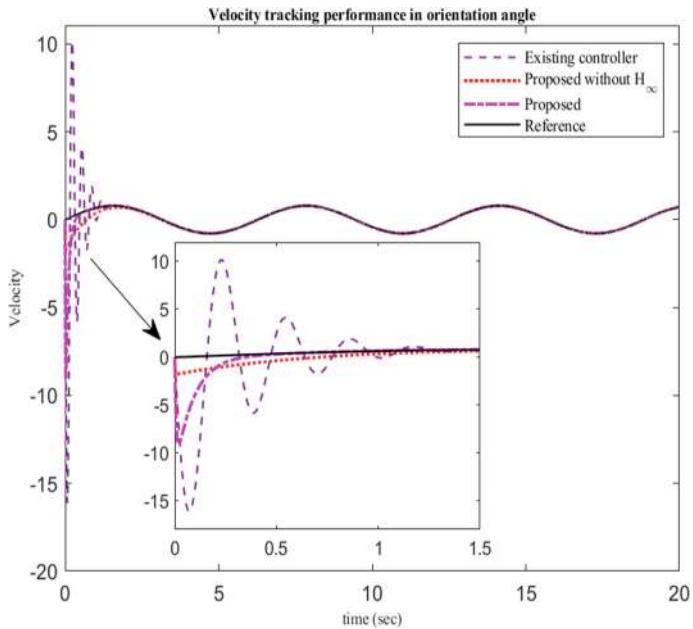
**Fig. 3** Position in y-direction



**Fig. 4** Position in orientation angle



**Fig. 5** Velocity in y-direction



**Fig. 6** Velocity in orientation angle

**Table 1**  $L^2$ -norm of position-tracking error

Controllers	$L^2[\bar{\chi}_1]$	$L^2[\bar{\chi}_2]$
Existing controller	0.0981	0.0943
Proposed controller without $H_\infty$	0.1679	0.1671
Proposed controller	0.0723	0.0712

show the supremacy of the designed control approach. The velocity tracking performance of the nonholonomically constrained mobile robot system is displayed in Figs. 5 and 6, which show the enhanced performance of the designed control technique. However, over a few periods, the system becomes quite smooth in following the reference velocity with the designed approach comparatively. So these results show that the designed control technique tracks the desired trajectory robustly and precisely in an efficient manner.

Additionally,  $L^2$  norm error analysis is provided in tabular form by comparing the outcomes with the existing controller given in Table 1. A reduced tracking error is shown by a lower value of  $L^2[\bar{\chi}]$ , indicating the efficacy of the control approach.

## 6 Conclusions

This work examines the trajectory tracking control problem of a nonholonomic mobile robot under external disturbances and uncertainty in the system. The design control technique utilized a fractional-order sliding surface for the faster response of the controller. To manage outside disruptions and uncertainties in a robust manner, an  $H_\infty$ -based robust fractional-order SMC technique is presented. The model-based approach is used in the controller's design with  $H_\infty$  tracking control to offset the adverse impacts of the system uncertainties and disruptions. The Lyapunov approach is successfully implemented to analyze the stability of the dynamical system. In the last, a numerical simulation study is compelled in a comparative way that shows the supremacy of the designed control technique. So from the overall analysis, it can be stated that the designed control scheme is efficient and handles the outside interfaces appropriately.

## References

1. Wu X, Wang Y, Dang X (2014) Robust adaptive sliding-mode control of condenser-cleaning mobile manipulator using fuzzy wavelet neural network. *Fuzzy Sets Syst* 235:62–82
2. Xie H, Zheng J, Chai R, Nguyen HT (2021) Robust tracking control of a differential drive wheeled mobile robot using fast nonsingular terminal sliding mode. *Comput Electr Eng* 96:107488

3. Dong W, Huo W (1999) Tracking control of wheeled mobile robots with unknown dynamics. In: Proceedings 1999 IEEE International conference on robotics and automation (Cat. No. 99CH36288C), vol 4. IEEE, pp 2645–2650
4. Chen N, Song F, Li G, Sun X, Ai C (2013) An adaptive sliding mode backstepping control for the mobile manipulator with nonholonomic constraints. *Commun Nonlin Sci Numer Simul* 18(10):2885–2899
5. Beginini M, Bertol DW, Martins NA (2017) A robust adaptive fuzzy variable structure tracking control for the wheeled mobile robot: simulation and experimental results. *Control Eng Pract* 64:27–43
6. Chaudhary KS, Kumar N (2023) Fractional order fast terminal sliding mode control scheme for tracking control of robot manipulators. *ISA Trans* 142:57–69
7. Kumar N, Chaudhary KS (2024) Neural network based fractional order sliding mode tracking control of nonholonomic mobile robots. *J Comput Anal Appl* 33(1)
8. Xie Y, Zhang X, Meng W, Zheng S, Jiang L, Meng J, Wang S (2021) Coupled fractional-order sliding mode control and obstacle avoidance of a four-wheeled steerable mobile robot. *ISA Trans* 108:282–294
9. Kumar N, Chaudhary KS (2024) Motion control of underactuated cart-double-pendulum system via fractional-order sliding mode controller. In: International conference on soft computing: theories and applications, pp 155–165
10. Eray O, Tokat S (2020) The design of a fractional-order sliding mode controller with a time-varying sliding surface. *Trans Inst Meas Control* 42(16):3196–3215
11. Panwar V (2017) Wavelet neural network-based h-infinity trajectory tracking for robot manipulators using fast terminal sliding mode control. *Robotica* 35(7):1488–1503

# MediSentBot—Medicine Review Sentiment Analysis and Recommendation Bot Using Modern NLP



Shreya Rajpal and E. S. Madhan

**Abstract** Numerous studies have reported that adverse drug events (ADEs) occur in 5 to 35 percent of patients across all age groups in outpatient settings. With developments in artificial intelligence (AI) and Natural Language Processing (NLP), the authors have employed NLP techniques to design a system that aims to mitigate adverse drug reaction (ADR) levels and facilitate their early identification. This paper presents an analysis and comparative study of traditional and modern techniques for sentiment analysis, focusing on their applicability within the healthcare landscape, using the UCI ML-Drug Review Dataset. Traditional methods, such as Long Short-Term Memory (LSTM) networks, Random Forest, and LightGBM, are presented against advanced language models (LMs) like LLaMA 2-7B, GPT-3.5 Turbo, and Gemma 7B. The experiments demonstrate that LMs, particularly GPT-3.5 Turbo in a 15-shot setting, outperform traditional methods with an impressive accuracy of 94.20%. Based on these findings, the authors developed MediSentBot, a medicine recommendation and review system utilising the GPT-3.5 Turbo model to recommend medicines and aggregate live reviews, ensuring up-to-date and accurate sentiment analysis. This study highlights the enhanced performance and applicability of modern NLP techniques in medical downstream tasks, specifically sentiment analysis.

**Keywords** Language models · LSTM · Neural networks · Natural language processing · Medical analysis

---

S. Rajpal · E. S. Madhan  
Vellore Institute of Technology, Vellore, Tamil Nadu, India  
e-mail: [shreya.rajpal2021@vitstudent.ac.in](mailto:shreya.rajpal2021@vitstudent.ac.in)

E. S. Madhan  
e-mail: [madhan.es@vit.ac.in](mailto:madhan.es@vit.ac.in)

## 1 Introduction

Adverse drug reactions (ADRs) have been a big concern for the health sector since they contribute to morbidity and mortality on a global scale. According to a study, one out of six hospitalised patients aged 65 years or older suffers from major adverse drug reactions [1]. Most adverse drug reactions (ADRs) remain undetected prior to the practical use of medication, even after comprehensive testing during clinical trials. This is particularly true for those populations not considered in clinical trials, such as children, the elderly, or people with multiple conditions [1, 2]. A study projected that as many as 3% of all patients suffer preventable harm from medications, further underlining the critical nature of good pharmacovigilance systems [3].

The issue can be addressed by developing a system that analyzes positive and negative reviews of medicinal drugs over a period of time. The lack of such information leads to treatment decisions that may not be ideal, worsening patient outcomes and increasing the cost of healthcare. This challenge serves as the motivation for conducting this research, with the objective to explore innovative solutions to enhance the quality of drug-related information and ultimately improve healthcare outcomes. Natural Language Processing (NLP) has become an integral tool for making meaningful insights from extensive datasets containing unstructured text data, like patient reviews or clinical notes. Sentiment analysis has used traditional models like Random Forest and LightGBM, and deep learning models such as Long Short-Term Memory (LSTM). However, these approaches often struggle to capture complex linguistic nuances and long-range dependencies compared to transformer-based architectures.

A major development in NLP techniques are language models (LMs). These include open-source and closed-source models, such as, Llama, GPT, Gemini, and Claude [4]. With significant pre-training, fine-tuning, and deep learning architectures, these models are able to comprehend and produce text with high accuracy. They also show significant progress in different downstream NLP tasks. They improve sentiment analysis accuracy by being able to manage a wide range of inputs and capture contextual nuances [4].

The dataset used in this research was retrieved from the UCI Machine Learning Repository [5]. It contains patient reviews of different drugs, a brief description of the condition for which that drug is taken, and a user rating on a scale of 1 to 10. The dataset is characterised by six features: A drug name, a condition for which the drug is prescribed, a patient's narrative review, the satisfaction rating given to it, the submission date, and the number of users who found the respective review helpful.

Developing a review-based medicinal drug review system would revolutionise pharmacovigilance; the system could use state-of-the-art technology and constantly update the information. Such systems can aggregate real-life reviews and, therefore, provide current and comprehensive information with which patients, regulatory bodies, and clinicians can make informed decisions. This proactive system will lead to better care for patients and boost surveillance over drug-related issues.

In this research, the authors developed a medicinal drug recommendation and review system named MediSentBot by utilising the 15-shot GPT-3.5 Turbo model to apply sentiment analysis on medicinal drug reviews. MediSentBot continuously aggregates real-time reviews for the refinement of its dataset, and gives recommendations on medications considering positive user feedback. The system developed here provides new powerful tools that can be used to improve decision-making for healthcare, demonstrating the potential of modern NLP techniques to enhance pharmaceutical surveillance and pharmacovigilance. The demo and code is available at [www.medisentbot.projectsbyshr.in](http://www.medisentbot.projectsbyshr.in).

## 2 Literature Review

Traditional sentiment analysis techniques in the domain of drug reviews have primarily focused on using various machine learning algorithms and feature engineering methods to classify and extract sentiments from textual data. Vijayaraghavan and Basu [6] applied some machine learning models, including count vectors (CV), and term frequency-inverse document frequency (TF-IDF) as supervised techniques to carry out a sentiment analysis study on medication reviews. They trained their models on UCI Machine Learning Repository datasets to predict the sentiment category of drug ratings from textual critiques. This worked well, particularly for labels such as “birth control”, “depression”, and “pain”.

Uddin et al. [7] conducted similar research by evaluating a corpus of medication reviews using classifiers such as Naive Bayes classifiers, Random Forests, Support Vector Classifiers (SVCs), and Multilayer Perceptron. They found that the Random Forest was the best-performing algorithm in this class of algorithms. Additionally, they applied linear SVC for multiclass classification, and performance was effective.

With developments in the field of NLP, there have been more sophisticated models utilising deep learning and language models. Punith and Raketla [8] conducted a study where they compared transformer-based models like BERT and XLNET to neural network-based models with ELMO word embeddings. For instance, their results show that transformer models work significantly better than conventional techniques in terms of accuracy and the power to capture complex linguistic nuances pertaining to medication reviews.

In a study by Rathod et al. [9] on machine learning methods for the sentiment analysis in pharmaceutical reviews; the authors studied how different ways of feature engineering—for instance, bag-of-words, n-grammes, or word embeddings—affect the performance of these models. Their study indicated that strategic application of such techniques, with advanced machine learning algorithms, can lead to high levels of accuracy and F1-scores, thereby, attempting to extract useful insights from unstructured drug review data.

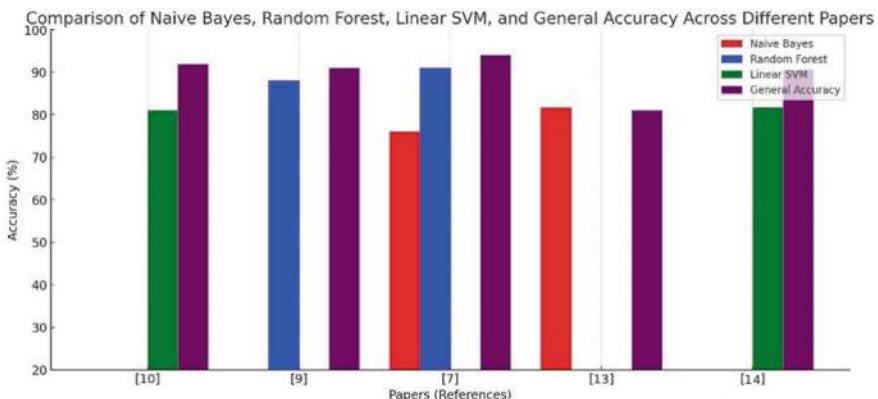
Phatak et al. performed a comprehensive study of sentiment analysis for drug reviews using user opinions [10]. They presented a multi-step process that involves mining product features, locating opinion sentences, and summarizing the results to

analyze customer reviews. It differs from generic text summarization by focusing specifically on the product aspects users comment on, and whether those comments are positive or negative.

Ruiz and Bedmar [11] present an overview of the performance of hybrid deep learning classifiers, especially the combination of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks and BERT combined with a bi-directional LSTM. The results show that though CNN models are of acceptable effectiveness with lower training time, BERT in combination with bi-directional LSTM slightly outperforms other models, confirming the possibility for a model to learn these nuanced linguistic features necessary in understanding the user-generated content of drug reviews.

Korkontzelos et al. [12] studied the use of sentiment analysis tools for the discovery of mentions of adverse drug reactions in social media and health forums. The research focused on improving activities of pharmacovigilance by detecting ADR mentions with more precision, hence supported by large-scale expression of public opinion available on these platforms. This shows the paramount significance of sentiment analysis in improving the efficiency of pharmacovigilance systems and safety of drugs (Fig. 1).

The literature on sentiment analysis of medication reviews shows a definite shift from conventional machine learning approaches to sophisticated NLP techniques involving LMs. Traditional methods, while useful, may fall short of handling complex linguistic nuances, for instance, sarcasm, compared to modern transformer-based language models. Sentiment analysis has a lot of potential to improve drug safety monitoring and offer insightful data on patient experiences and drug efficacy when integrated into pharmaceutical monitoring systems. The creation of sophisticated systems like MediSentBot, which make use of cutting-edge Natural Language Processing models, is a major advancement in the application of sentiment analysis to improve healthcare decision-making.



**Fig. 1** Comparison of Naïve Bayes, Random Forest, linear SVM, highest accuracy

### **3 Methodology**

### 3.1 Dataset

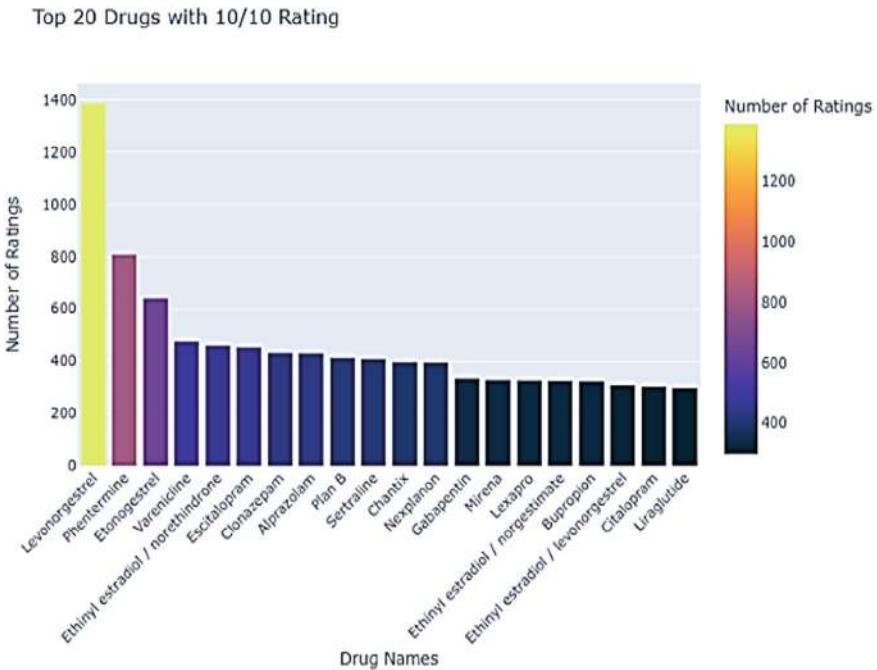
The dataset utilised is the Drug Review Dataset from the UCI Machine Learning Repository, comprising patient reviews on specific drugs, related conditions, and ratings reflecting overall patient satisfaction. The sentiment of the reviews, derived from ratings, is categorised as positive, neutral, or negative. The authors perform exploratory data analysis on the dataset [5] and gain the following insights (Figs. 2, 3, 4, 5 and 6).



**Fig. 2** Word cloud of reviews



**Fig. 3** Word cloud of positive words



**Fig. 4** Top 20 drugs with most positive rating in the dataset

The preprocessing steps for the dataset included:

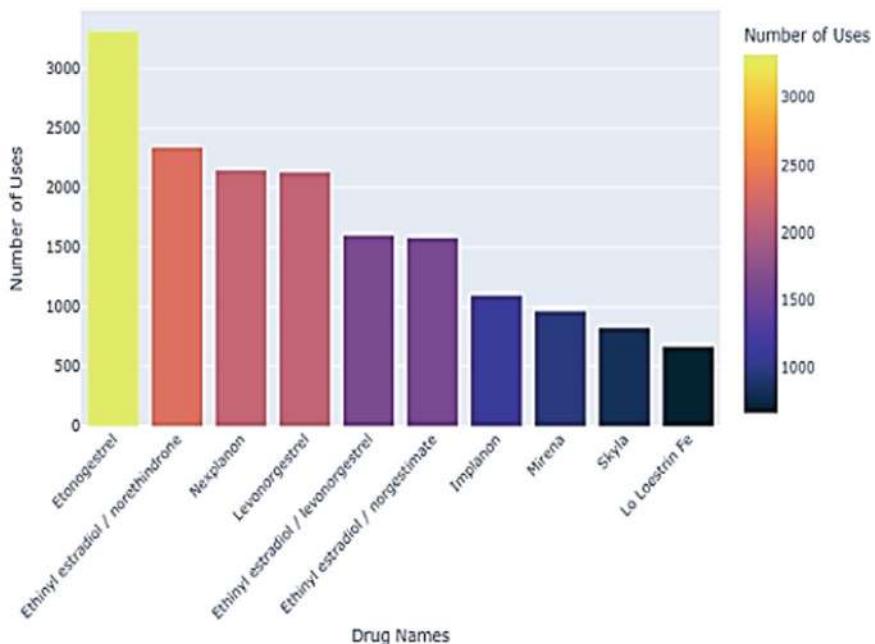
- Cleaning: Removal of special characters and normalisation of numerical values.
- Tokenisation: Conversion of text into sequences of integers.
- Padding: Ensuring uniform sequence length for model input.
- Feature Extraction: Transforming reviews into numerical features using TF-IDF or word embeddings.

### 3.2 Traditional NLP Methods

#### 3.2.1 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are a type of Recurrent Neural Network (RNN) designed to capture long-term dependencies in sequence data through gating mechanisms that regulate information flow, mitigating the vanishing gradient problem [15, 16]. The LSTM model architecture included the embedding layer, Convolutional Layer, MaxPooling Layer, the LSTM Layer that captures temporal dependencies in the data, and the Dense Layer with a softmax activation function [15]. The train and test splits for all the methods were in a 70:30 ratio.

### Top 10 Medicine Drugs used for Birth Control



**Fig. 5** Top 10 medicinal drugs used for birth control from the dataset, based on reviews the developed system predicts the most efficient medicinal drug for birth control

Mathematically, LSTM consists of:

Forget Gate: Decides what information to discard from the cell state.

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

Input Gate: Decides which values to update.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

Cell State Update: Updates the cell state with new information.

$$\overline{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * \overline{C}_t \quad (4)$$

Output Gate: Decides the output based on the cell state.

Top 20 Conditions in the dataset by count

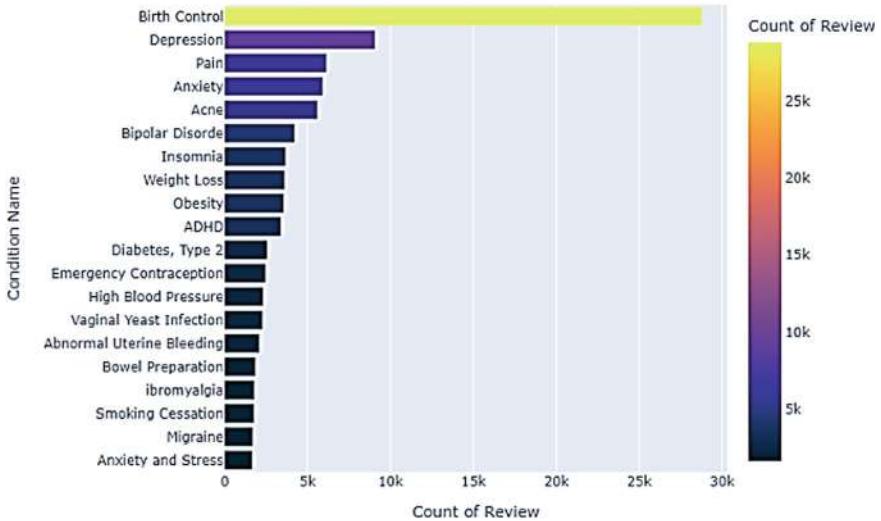


Fig. 6 Top 20 conditions in the dataset by count in the dataset

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

### 3.2.2 Random Forest

The Random Forest algorithm improves predictive accuracy and mitigates overfitting through the aggregation of multiple decision trees, each trained on distinct data subsets [17].

Mathematically, Random Forest consists of:

- Aggregation: The final prediction  $\hat{y}$  is made by majority voting among all the trees' predictions for classification tasks.

$$\hat{y} = \text{mode}(\{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n\}) \quad (6)$$

- For regression tasks, the final prediction is calculated as the average of all trees' predictions.

$$\hat{y} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i \quad (7)$$

The model contains a series of steps like Feature Extraction: Text reviews were transformed into numerical features using techniques such as TF-IDF or word embedding, a Random Forest classifier was instantiated with 100 trees and trained using the training dataset and cross validation. This can also be done by addressing the high variance problem of decision trees by using bootstrapping and random feature selection. However, the paper's focus is on preventing overfitting, hence this method is not considered [7].

### **3.2.3 Light Gradient Boosting Machine**

Light Gradient Boosting Machine (LightGBM) is a gradient boosting framework that uses histogram-based algorithms to find optimal splits, making it significantly faster than traditional gradient boosting methods. This was used to build the classification model due to its efficiency, speed, and accuracy, making it suitable for large datasets. The features selected for the model included condition, sentiment, day, etc. [18, 19].

The objective function for binary classification can be written as:

$$L(y, \hat{y}) = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (8)$$

where  $y_i$  is the true label and  $\hat{y}_i$  is the predicted probability.

The feature importance plot from the LightGBM model indicated that the most significant features were mean word length and condition, while the least significant feature was the upper-case word count.

## **3.3 Modern NLP: Language Model Implementation and Evaluation**

### **3.3.1 Data Preprocessing**

Data preprocessing is a crucial step in NLP model development. For their initial evaluation study, the authors used the Keras tokeniser, but for the language model study, Latex content from review texts was removed to begin processing the reviews followed by a cleaning pipeline [20] [Appendix 9.5].

### **3.3.2 Zero-Shot and Few-Shot Prompting**

Three language models were prompted in a zero-shot manner and their pre-trained knowledge was leveraged to evaluate the texts for a specific downstream task. For few-shot prompting, the authors considered two samples, a 3 shot and a 15 shot;

this difference was chosen to maximise the variability of the texts the model can receive from the training set of the dataset. These models were prompted to perform sentiment analysis designed to classify reviews as positive (corresponding to ratings 8–10), neutral, or negative (corresponding to ratings 0–3) [Appendix 9.5]. The prompt templates are available in the appendix.

The authors used state-of-the-art models like Llama 2-7B, GPT3.5, and Gemma 7B for the 0 shot and few shot analyses in this paper [21–23]. Llama 2-7B, an open-source language model by Meta, is designed for a wide range of natural language understanding tasks and employs Grouped-Query Attention (GQA) to enhance performance and scalability [24]. GPT3.5 is also one of the most powerful language models. Gemma 7b is an open-source language model developed by Google. In this paper, the authors discuss how these models perform on the downstream task of sentiment analysis on medicine review data.

### 3.3.3 Open Versus Closed-Source Models

There is also a discussion in the AI community regarding open vs. closed-source language models. In their research, the authors incorporate both to gain a better understanding of the language model’s performance in downstream tasks in the medical domain. Models like LlaMA 2-7B are open source, providing transparency and accessibility. Closed-source models, such as the GPT-3.5 Turbo, offer higher performance and advanced safety features but limit customisation and access.

## 4 Results and Discussion

### 4.1 Machine Learning Techniques

The LightGBM model outperformed both the LSTM and Random Forest models, achieving the highest accuracy of 90.15%. The LSTM model also demonstrated strong performance with an accuracy of 86.02%, highlighting its ability to capture temporal dependencies in text data. The Random Forest model, while useful, had the lowest accuracy at 62.00%, indicating that ensemble methods may not be as effective as deep learning approaches or gradient boosting in this context. Given the results, the LightGBM model is recommended for its superior performance and efficiency in handling large datasets. However, the LSTM model is also a viable option when capturing sequential dependencies is crucial. Random Forest, while less accurate in this case, can still be valuable for its simplicity and interpretability (Table 1).

**Table 1** Accuracy depiction of traditional NLP models

Model	Accuracy (%)
LSTM	86.02
Random Forest	62.00
LightGBM	90.15

## 4.2 Language Models (LMs)

*LlaMA 2-7B*: With a 65.33% accuracy percentage, the model performed second-best for 0 shots. However, the results for 15 shot evaluation were inconclusive as the model gave no or irrelevant and incoherent responses. Therefore, it became impractical for extensive testing (Table 2).

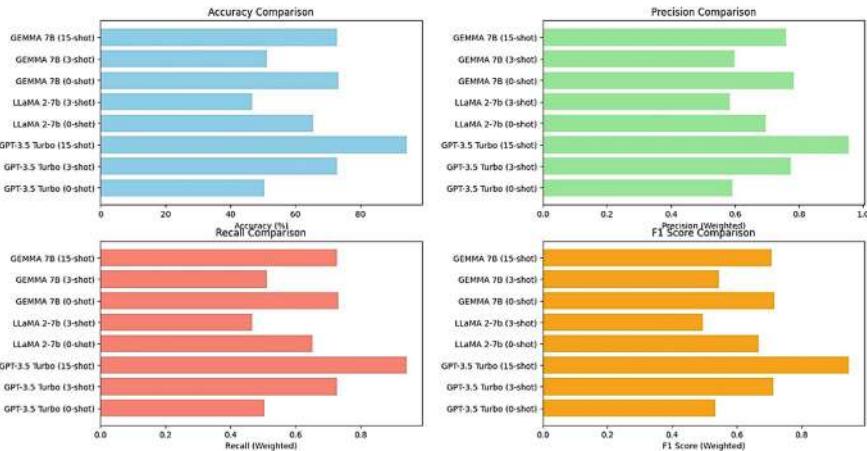
*GPT-3.5 Turbo*: While GPT-3.5 Turbo delivered excellent performance, particularly in the 15-shot prompting scenario with an accuracy of 94.20% and in the 3-shot performance with an accuracy of 72.80%, it is a closed-source model. It raises concerns about potential issues like hallucinations, bias, and a lack of transparency. Since users cannot inspect or modify the underlying model, they are exposed to risks that could affect the reliability and fairness of the outputs (Fig. 7).

*Gemma 7B*: Gemma 7B, a new open-source model, showed promising results. In zero-shot scenarios, it achieved an accuracy of 73.20%, demonstrating strong performance. It was the second-best language model for 3-shot prompts with an accuracy of 51.20% and for 15-shot prompts with an accuracy of 72.80%. Gemma 7B's open-source nature allows for greater flexibility and transparency, enabling users to modify and adapt the model as needed, reducing the risk of hidden biases, and improving overall trustworthiness.

*Model Performance*: The LightGBM model emerged as the top performer among traditional machine learning models with an accuracy of 90.15%, followed closely by the GPT-3.5 Turbo with 94.20% in the 15-shot scenario. LSTM models also performed well, capturing sequential dependencies effectively with an accuracy of 86.02%.

**Table 2** Computational results of GPT 3.5 Turbo, LLaMA 2-7B, and Gemma 7B, [Bold: Second-best results], [yellow highlight: Best result in that prompt type], [blue highlight: Overall best result]

Model	Prompt Type	Accuracy	Precision (Weighted)	Recall (Weighted)	F1 Score (Weighted)
GPT-3.5 Turbo	Zero-shot	50.40%	0.591	0.504	0.533
GPT-3.5 Turbo	3-shot	<b>72.80%</b>	<b>0.774</b>	<b>0.728</b>	<b>0.713</b>
GPT-3.5 Turbo	15-shot	<b>94.20%</b>	<b>0.956</b>	<b>0.942</b>	<b>0.948</b>
LLaMA 2-7B	Zero-shot	<b>65.33%</b>	<b>0.696</b>	<b>0.653</b>	<b>0.668</b>
LLaMA 2-7B	3-shot	46.60%	0.584	0.466	0.495
LLaMA 2-7B	15-shot	-	-	-	-
Gemma 7B	Zero-shot	<b>73.20%</b>	<b>0.784</b>	<b>0.732</b>	<b>0.717</b>
Gemma 7B	3-shot	<b>51.20%</b>	<b>0.599</b>	<b>0.512</b>	<b>0.545</b>
Gemma 7B	15-shot	<b>72.80%</b>	<b>0.759</b>	<b>0.728</b>	<b>0.708</b>



**Fig. 7** Graphical representation of their computational results

**Recommendations:** Given the results, the LightGBM model in traditional NLP and GPT-3.5 Turbo in modern NLP are highly effective (especially GPT with few-shot prompting, though users should be cautious of its closed-source nature). Gemma 7B is a strong contender in the open-source space, providing a good balance of performance and flexibility. The authors also prevent overfitting by incorporating a mix of genuine reviews from the training dataset and a substantial proportion of GPT-generated reviews. These generated reviews are further annotated by humans for sentiment analysis.

**Sentiment Analysis and Modern NLP:** Modern NLP techniques, powered by advanced models like GPT-3.5 Turbo, LLaMA 2-7B, and Gemma 7B, offer significantly improved performance over traditional methods in tasks such as sentiment analysis. These models can understand context better, handle sarcasm and nuanced language, and provide more accurate and reliable results. Their ability to perform zero-shot and few-shot learning enables them to generalise well from minimal data, reducing the need for extensive labelled datasets. This makes them superior to older NLP techniques, which often required extensive feature engineering and struggled with complex language nuances.

#### 4.3 Medicine Recommendation System

Based on the research, the authors developed a system that processes live customer reviews about different medicines, thoroughly performing sentiment analysis using GPT-3.5 with 15-shot capabilities. This system formulates the best medicinal drug recommendations for each condition. On one end, users can access a detailed analysis of medicinal drug performance for specific conditions, allowing them to make

<b>Algorithm 1 MediSentBot: Medicine Sentiment Analysis and Recommendation Bot</b>	
<b>Input:</b>	Live customer reviews about different drugs
<b>Output:</b>	Best medicinal drug recommendation for each condition
<b>Step 1: Data Collection</b>	Collect live reviews from customers about different medicinal drugs.
<b>Step 2: Data Preprocessing</b>	Perform sentiment analysis using GPT-3.5 with 15-shot capabilities to classify reviews.
<b>Step 3: Filtering Positive Reviews</b>	Filter out neutral reviews to avoid recommendations of medicinal drugs with tolerable or present side effects.
<b>Step 4: Medicinal Drug Filtering</b>	Filter the top-rated drugs based on positive reviews to ensure only the most highly recommended drugs are considered.
<b>Step 5: Medicinal Drug Recommendation System (MediSentBot)</b>	Develop MediSentBot as a chat model using GPT-3.5 to interact with patients.
<b>Step 6: Medicinal Drug Recommendation System (MediSentBot)</b>	MediSentBot discusses health, symptoms, and doctor recommendations with patients.
<b>Step 7: Continuous Dataset Expansion</b>	Based on the sentiment analysis, MediSentBot suggests the medicinal drug with the most positive reviews.
	Continuously expand the dataset by incorporating live reviews.

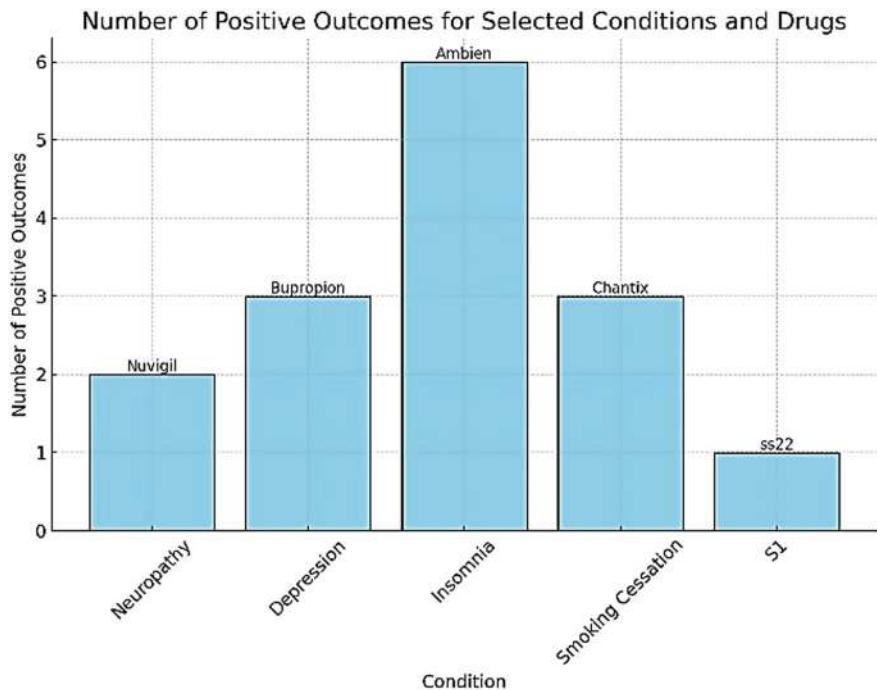
**Fig. 8** Algorithm of MediSentBot: Medicine recommendation system. MediSentBot is available for access at [www.medisentbot.projectsbyshr.in](http://www.medisentbot.projectsbyshr.in)

informed decisions based on comprehensive sentiment analysis. On the other hand, the authors have used GPT-3.5 as a conversational model for MediSentBot, a bot that interacts with patients about their health, symptoms, and doctor recommendations. MediSentBot then suggests the medicinal drug with the most positive reviews based on the review data (Figs. 8 and 9).

By including only positive reviews, the system effectively filters out any medicines that may have worked well but had side effects that can be classified as a neutral review, thereby avoiding them. This approach helps prevent side effects from reaching customers. Additionally, the system is designed to continuously expand its dataset by incorporating live reviews, ensuring it remains up-to-date and accurate. This dual-end approach not only provides valuable insights into medicinal drug performance but also enhances patient interaction and medicinal drug recommendation accuracy.

## 5 Disclaimer

MediSentBot is a research prototype designed to analyze patient-review sentiment and is not intended to provide medical advice without the presence of a doctor, diagnosis, or treatment recommendations. Outputs should not be used as a substitute for consultation with a qualified healthcare professional. Users must comply with



**Fig. 9** Bar graph of some data from live and dataset reviews by MediSentBot, here for testing purpose ss22 is a medicinal drug introduced in the system as a live review which was positive by the customer

all applicable privacy and regulatory requirements (e.g., GDPR, HIPAA, FDA/CE regulations) when processing patient data.

## 6 Limitation

Due to the hardware constraints, running LLaMA 2-7B for 15-shot prompts on a sample set of 100 reviews was time-consuming. Additionally, the results for the 10 reviews processed were inconclusive and included repetitive texts from the prompts. Despite the authors' efforts to include three different LMs, future research could benefit from exploring how medical sentiment analysis models fine-tuned for specific tasks might perform better. It is also important to note that the use of language models often brings inherent concerns related to fairness and bias in their responses.

## 7 Conclusion

In this study, the authors conducted sentiment analysis using both traditional and modern Natural Language Processing (NLP) approaches to evaluate the performance and accuracy of various models. Based on these findings, while traditional and state-of-the-art NLP methods show comparable performance to GPT-3.5 under a 15-shot setting, GPT-3.5 consistently demonstrates significantly superior results overall. The Gemma model demonstrated a high value of performance, and LightGBM model also resulted in good accuracy. This comparison brings to the surface the debate there has been about open-source and closed-source models and their related transparency issues.

Based on these insights, the authors developed MediSentBot, which used the advanced capabilities of GPT-3.5 for sentiment analysis and medicinal drug recommendation. The application thus puts forth a full solution for the automatic analysis of customer reviews and recommendations for medicinal drugs, in a personalised manner, based on the positive feedback. Future work will look towards exploring zero-shot learning capabilities with Gemma to take more advanced directions to further push the abilities of the system. This will ensure that in the long run, MediSentBot provides the right, reliable and cost effective recommendations while having a robust and expandable dataset.

## References

1. World Health Organisation (2018) Safety of medicines—adverse drug reactions
2. BMC Medicine (2023) Preventable medication harm across health care settings: a systematic review and meta-analysis
3. Jennings ELM, Murphy KD, Gallagher D, O’Mahony D (2020) In-hospital drug reactions in older adults; prevalence, presentation and associated medicines—a systematic review and meta-analysis. *Age and Ageing*. <https://academic.oup.com/ageing/article/49/6/948/5918299>
4. Zhang W, Deng Y, Liu B, Pan SJ, Bing L (2023) Sentiment analysis in the era of large language models: a reality check. In: DAMO Academy, Alibaba Group, Nanyang Technological University, Singapore, University of Illinois at Chicago, The Chinese University of Hong Kong. <https://arxiv.org/pdf/2305.15005>
5. UCI Machine Learning Repository (2024) Drug review sentiment analysis using boosting algorithms. <https://www.kaggle.com/datasets/jessicali9530/kuc-hackathon-winter-2018>. Last accessed: 20 July 2024
6. Vijayaraghavan S, Basu D (2020) Sentiment analysis in drug reviews using supervised machine learning algorithms. <https://arxiv.org/pdf/2003.11643>
7. Uddin MN, Hafiz MFB, Hossain S, Islam SMM (2022) Drug sentiment analysis using machine learning classifiers. Department of Computer Science and Engineering, East Delta University, Chattogram, Bangladesh. [https://thesai.org/Downloads/Volume13No1/Paper\\_12-Drug\\_Sentiment\\_Analysis\\_using\\_Machine\\_Learning.pdf](https://thesai.org/Downloads/Volume13No1/Paper_12-Drug_Sentiment_Analysis_using_Machine_Learning.pdf)
8. Punith NS, Raketa K (2021) Sentiment analysis of drug reviews using transfer learning. In: 2021 Third international conference on inventive research in computing applications (ICIRCA). IEEE, pp 1–6. <https://ieeexplore.ieee.org/document/9544574>

9. Rathod D, Patel K, Goswami AJ, Degadwala S, Vyas D (2023) Exploring medicinal drug sentiment analysis with machine learning techniques. IEEE. <https://ieeexplore.ieee.org/document/10134055>
10. Phatak S, Patil SV, Patil R, Anekar N (2021) A review on: sentiment analysis for medicinal drug using user reviews. Annasaheb Dange College of Engineering, Sangli, India. [https://www.researchgate.net/publication/304996774\\_A REVIEW ON SENTIMENT ANALYSIS FOR DRUG USING USER REVIEWS](https://www.researchgate.net/publication/304996774_A REVIEW ON SENTIMENT ANALYSIS FOR DRUG USING USER REVIEWS)
11. Colón-Ruiz C, Segura-Bedmar I (2020) Comparing deep learning architectures for sentiment analysis on drug reviews. J Biomed Inform 110:103539. <https://doi.org/10.1016/j.jbi.2020.103539>
12. Korkontzelos I, Nikfarjam A, Shardlow M, Sarker A, Ananiadou S, Gonzalez GH (2016) Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts. <https://pubmed.ncbi.nlm.nih.gov/27363901/>
13. Na J, Kyaing WYM (2015) Sentiment analysis of user-generated content on drug review websites. J Inf Sci Theory Pract 3(1):6–23. <https://doi.org/10.1633/JISTaP.2015.3.1.1>
14. Mahajan, A., Ray, A., Verma, A., Kohad, S., & Thakare, P. N. (2021). Sentiment Analysis using Supervised Machine Learning. International Journal of Advanced Research in Computer Science and Software Engineering, 11(5), 102–108. [https://ijariie.com/AdminUploadPdf/Sentiment\\_Analysis\\_Using\\_Supervised\\_Machine\\_Learning\\_ijariie13051.pdf?srsltid=AfmBOo06fpYTJ2q8tkViZ25t6rwY8vIEl\\_1Y9DlOUIDWBbb0FJ1o9A](https://ijariie.com/AdminUploadPdf/Sentiment_Analysis_Using_Supervised_Machine_Learning_ijariie13051.pdf?srsltid=AfmBOo06fpYTJ2q8tkViZ25t6rwY8vIEl_1Y9DlOUIDWBbb0FJ1o9A)
15. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
16. GeeksforGeeks Homepage (2024) Deep learning: introduction to long short-term memory. <https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/>. Last accessed 02 July 2024
17. Jihad A, Khan R, Ahmad N, Maqsood I (2012) Random forests and decision trees. Int J Comput Sci Issues (9)
18. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y (2017) LightGBM: a highly efficient gradient boosting decision tree. In: Guyon I, Von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R (eds) Advances in neural information processing systems, vol 30. Curran Associates, Inc. [https://proceedings.neurips.cc/paper\\_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf)
19. Mishra S (2021) Drug review sentiment analysis using boosting algorithms. Electrical and Electronics Department, Bharati Vidyapeeth's College of Engineering, New Delhi, India
20. Kumawat S, Yadav I, Pahal N, Goel D (2021) Sentiment analysis using language models: a study. In: 2021 11th international conference on cloud computing, data science & engineering (confluence), Noida, India, pp 984–988. <https://doi.org/10.1109/Confluence51648.2021.9377043>
21. Team G, Mesnard T, Hardin C, Dadashi R, Bhupatiraju S, Pathak S et al (2024) Gemma: open models based on gemini research and technology. [arXiv:2403.08295](https://arxiv.org/abs/2403.08295)
22. Touvron H, Lavril T, Izacard G, Martinet X, Lachaux MA, Lacroix T, Rozière B, Goyal N, Hambré E, Azhar F, Rodriguez A, Joulin A, Grave E, Lample G (2023) LLaMA: open and efficient foundation language models. [arXiv:2302.13971](https://arxiv.org/abs/2302.13971)
23. Ye J, Chen X, Xu N, Zu C, Shao Z, Liu S, Cui Y, Zhou Z, Gong C, Shen Y, Zhou J, Chen S, Gui T, Zhang Q, Huang X (2023) A comprehensive capability analysis of GPT-3 and GPT-3.5 series models. [arXiv:2303.10420](https://arxiv.org/abs/2303.10420)
24. Ainslie J, Lee-Thorp J, de Jong M, Zemlyanskiy Y, Lebrón F, Sanghai S (2023) GQA: training generalised multi-query transformer models from multi-head checkpoints. Google Research. <https://arxiv.org/abs/2305.13245>

# Cybercrime Classification and Tracking Computations System Using Machine Learning



**Ch. Rupa, Sree Vardhan Sunkara, Rakesh Kota, G. Thippa Reddy, and Pratik Vyas**

**Abstract** In the constantly emerging context of cyberspace, the escalating threat of cybercrime demands a nuanced understanding of classification techniques. This work conducts a comparative analysis of Support Vector Machine (SVM), XGBoost, Random Forest, and Naive Bayes for effectively classifying cyber threats. The investigation delves into an in-depth exploration of cybercrime categories—social engineering, identity theft, malware, and espionage—with a specific focus on state-wide cybercrime data in India. The dataset, meticulously sourced from authorized channels, namely Computer Emergency Response Team for International Negotiation (CERT-IN) and National Cybercrime Reporting Portal (NCRP), provides a comprehensive view of cybercrime trends across regions from 2020 to 2023. Rigorous analysis is ensured through the preparation of unstructured data, allowing for a granular examination of the unique traits and societal implications of cybercrimes. The study evaluates the performance of SVM in comparison to XGBoost, Random Forest, and Naive Bayes, offering insights into the strengths and limitations of each algorithm. The findings contribute significantly to the cybersecurity domain, guiding practitioners and researchers in making informed decisions for robust cyber threat identification and mitigation strategies tailored to the diverse cyber landscape in different Indian regions.

---

Ch. Rupa · S. V. Sunkara · R. Kota

Velagapudi Ramakrishna Siddhartha Engineering College, Kanuru, Vijayawada, India

G. T. Reddy

The College of Mathematics and Computer Science, Zhejiang University, Hangzhou, China

Division of Research and Development, Lovely Professional University, Phagwara, India

Center of Research Impact and Outcome, Chitkara University, Rajpura, Punjab, India

P. Vyas (✉)

Department of Computer Science, Nottingham Trent University, Nottingham, UK

e-mail: [pratik.vyas@ntu.ac.uk](mailto:pratik.vyas@ntu.ac.uk)

## 1 Introduction

In recent years, cybercrime has posed as an important threat to individuals, organizations, and society at large, with its diverse and constantly evolving nature presenting formidable challenges to cybersecurity professionals. As the digital landscape continues to expand and integrate into various aspects of daily life, the need for effective cybercrime detection and classification mechanisms becomes increasingly critical. Cybercrime encompasses an extensive variety of illicit activities, like phishing, ransomware attacks, data breaches, etc., that have devastating effects on people, businesses, and governments [1, 2]. In this context, accurate and timely classification of cyber threats is crucial for enabling proactive responses and mitigating potential damages.

In addition to its practical implications for cybersecurity operations, cybercrime classification serves as a vital tool for advancing our understanding of the complex dynamics underlying digital crime. By studying cyber threat data and patterns, researchers can understand the tactics, techniques, and procedures (TTPs) employed by cybercriminals, shedding light on their motivations and modus operandi. The categorization of cybercrimes also supports the recognition of similarities and relationships between different types of cyber threat(s), therefore developing integrated threat intelligence architectures and predictive analytics frameworks.

The highlighted system has great potential to correctly classify a wide range of cybercrime types, with specific attention on commonly detected crime types which includes malware, industrial espionage, identity theft, and social engineering for the years 2020–2023. This work attempts to advance an understanding of the strengths and weaknesses of different approaches by comparing the performance of Support Vector Machine (SVM), Random Forest Classifier (RFC), XGBoost, and Naive Bayes algorithms, thus facilitating the detection system-building process of reliable cybercrime detection systems. The system will provide value to cybersecurity investigators by revealing an generated classification of the unique traits and behaviors associated with each type of cybercrime, supporting the proactive prevention and investigation of the crime type. The emphasized system of cybersecurity crime detection types 2020–2023 provides a resource for detectives when examining current cyber threat(s), combining crime types, and with emerging crime types as cyberspace grows in steps to combine system sizes and various modded configurations for enhancing entry to exploitable assets.

## 2 Related Works

Toldinas et al. [3] have proposed a study on the emergent threat of cybercrime, especially malware attacks and its consequences. The ensemble-based classification technique of this study aims at improving performance for malware detection by multiple classifier combinations. First, it uses a stacked ensemble of Dense networks and

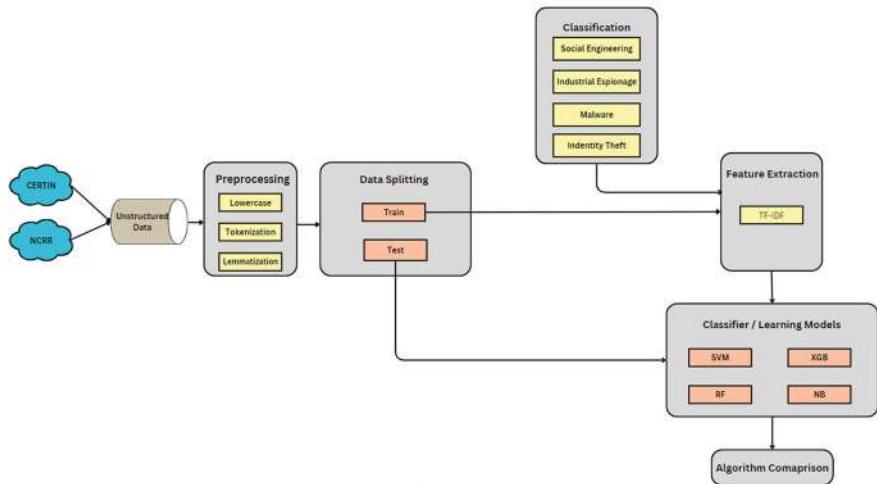
CNNs. Then the classification is fine-tuned by adding a meta-learner into the system. The second step is to explore various classifiers like KNN, SVM, and AdaBoost with respect to their performance as meta-learners to enhance the classification process. Finally, it measures the execution of the ensemble classification approach against the conventional approach to set a performance benchmark.

Djenna et al. in 2023 [4] proposed a study related to the rise in sophistication by cybercriminals and realistic cybersecurity measures that are becoming very essential in urgency. In turn, this work contributes to the betterment of cyber threat intelligence through improved understanding of the new forms of botnet attacks that are emerging. The work introduces a novel technique in identifying and detecting botnet attacks through utilizing unsupervised LSTM and CNN models. The approach will be validated using datasets CTU-13 and IoT-23 for checking the effectiveness of the approach in the detection of a botnet attack. Results show that the derived systems outperformed the others, attaining a positive rate of more than 98.7% with a FP rate of 0.04%.

Asam et al. [5] study emphasizes the critical role of malware analysis as the first line of defense against cyber-attacks, highlighting the importance of effective malware classification frameworks. In response, the authors introduce two novel malware classification methods: Deep Feature Space-based Malware classification (DFS-MC) and Deep-Boosted Feature Space-based Malware classification (DBFS-MC). In the DFS-MC method, deep features are extracted from modified CNN architectures and fed into SVM for malware classification. However, the DBFS-MC system creates boosted feature spaces by merging deep feature spaces from two tailored CNN architectures, which improves discrimination power. The study uses the hold-out cross-validation technique to assess the efficacy of the suggested classification frameworks, concentrating on malware variants that are challenging to distinguish, like Autorun. K.

### 3 Proposed Methodology

The proposed system delineates a comprehensive methodology for the comparative analysis of cybercrime classification utilizing ML algorithms. With a focus on evaluating the efficacy of SVM, Random Forests, XGBoost, and Naive Bayes across a spectrum of cybercrime categories, the study employs sophisticated preprocessing techniques including tokenization, lemmatization, and TF-IDF. Subsequently, the trained models undergo meticulous testing to gauge their performance against key metrics. Illustrated in Fig. 1, the systematic delineation of the methodology underscores a structured approach toward comparing machine learning algorithms in the realm of cybercrime classification. The proposed methodology unfolds through 6 essential phases, each contributing crucially to the overarching objective of enhancing cybercrime detection and prevention efforts. The phases involved in the proposed system are (Table 1):

**Fig. 1** Proposed methodology**Table 1** Comparison with existing works

References	Year	Dataset	Algorithm	Application	Advantages	Disadvantages
[3]	2021	ClaMP Dataset	KNN, SVM, AdaBoost	Windows PE Malware detection	Multi-stage classification	Single crime classification
[4]	2023	CTU-13, IoT-23	LSTM, CNN	Botnet detection classification	Superior performance, LFP	Single crime classification
[5]	2021	Multiple datasets	DFS-MC, DBFS-MC	Malware variant detection	Enhanced discrimination	Limited generalizability
[6]	2020	UNSW-NB15, CICIDS2017	LR, SVM, KNN, RF	Real-world relevance, enhanced detection	Overfitting risks	
[7]	2022	Multiple datasets	SVM, KNN, RFC	Cyber Stalking Detection	High accuracy	Dataset dependency, single crime
[8]	2019	Self-structured	RFC, OneR	Aggression cases classification	Relevance to specific demographic	Ethical considerations, less scope
Proposed	2024	Self-structured	SVM, RFC, Naïve Bayes, XGBoost	Year	State-wise cybercrime classification, analysis	Multiple crimes classifications, dataset independence

### 3.1 Data Collection

In this phase, unstructured data pertaining to various cybercrime incidents is gathered from authorized sources, including Computer Emergency Response Team for International Negotiation (CERT-IN), among others, spanning a period of 2020–2023. This curated dataset comprises approximately 500 records, each meticulously organized and annotated with eight key attributes: incident, offender, victim, harm, year, location, state, amount, and attack. These sources provide valuable insights into real-world cyber threats and incidents, encompassing a wide range of malicious activities such as social engineering, malware attacks, identity theft, and espionage. An overview of the proposed dataset is provided in Fig. 2.

### 3.2 Data Preprocessing

Following data collection, the raw data passes through preprocessing to refine it for subsequent analysis. This phase encompasses several crucial steps aimed at enhancing the quality and usability of the data.

Let  $D = (x_1, y_1), (x_2, y_2), \dots (x_n, y_n)$  where  $x_i$  represent text features,  $y_i$  represent label.

**Tokenization:** It has been utilized to break down the text data into words individually, facilitating further analysis by breaking down the text into its fundamental components [9].

**Stop Words Removal:** It works toward normalizing the text data by turning them into a single format that might also include tasks related to changing all texts into lowercase, removing punctuation, or correcting spelling/formatting variations [10].

Incident	Offender	Victim	Harm	Year	Location	State	Amount	Attack
1. Bank Account Information Theft	Criminal	Individual	Financial Loss	2022	Andhra Pradesh	Andhra	16 Lakhs	Social Engineering
2. Theft of Trade Secrets	Hacker	System Software	Loss of Intellectual Property	2022	Chennai	TamilNadu	50000	Industrial Espionage
3. Ransomware Attack	Hacker	System Software	Loss of Data	2022	Delhi	Delhi	25000	Malware
4. Credit Card Information Theft	Criminal	Individual	Financial Loss	2023	Mumbai	Maharashtra	3 Lakhs	Social Engineering
5. Phishing Scam	Criminal	Company	Loss of Data	2022	Bangalore	Karnataka	12 Lakhs	Social Engineering
6. Ransomware Attack	Hacker	System Software	Loss of Data	2022	Pune	Pune	-	Malware
7. Online Banking Credentials Theft	Criminal	Individual	Financial Loss	2021	Kolkata	Bengal	1.5 Lakhs	Social Engineering
8. Data Breach	Hacker	Company	Loss of Intellectual Property	2023	New Delhi	Delhi	-	Industrial Espionage
9. CEO Fraud	Criminal	Individual	Financial Loss	2022	Hyderabad	Telangana	8 Lakhs	Social Engineering
10. Phishing Attack	Hacker	System Software	Loss of Data	2023	New Delhi	Delhi	-	Malware
11. Online Identity Theft	Criminal	Individual	Financial Loss	2020	Bangalore	Karnataka	2 Lakhs	Identity theft
12. Email Spoofing	Criminal	Individual	Loss of Data	2020	Mumbai	Maharashtra	9 Lakhs	Social Engineering
13. Data Theft	Hacker	System Software	Loss of Data	2023	Kolkata	Bengal	-	Malware
14. Credit Card Fraud	Criminal	Individual	Financial Loss	2021	Bangalore	Karnataka	4 Lakhs	Social Engineering
15. Insider Information Leak	Hacker	Company	Loss of Intellectual Property	2020	New Delhi	Delhi	-	Industrial Espionage
16. CEO Impersonation	Criminal	Individual	Financial Loss	2022	Mumbai	Maharashtra	6 Lakhs	Social Engineering
17. Phishing Scam	Hacker	Software	Data Breach	2023	Chennai	TamilNadu	-	Malware
18. Online Identity Theft	Criminal	Individual	Financial Loss	2020	Pune	Pune	2 Lakhs	Identity theft
19. Business Email Compromise	Criminal	Individual	Loss of Data	2022	Hyderabad	Telangana	5 Lakhs	Social Engineering
20. Malicious Software Attack	Hacker	Software	Loss of Data	2020	Bangalore	Karnataka	-	Malware
21. Online Identity Theft	Criminal	Individual	Financial Loss	2023	Kolkata	Bengal	2.2 Lakhs	Identity theft
22. Insider Information Leak	Hacker	Company	Loss of Intellectual Property	2022	New Delhi	Delhi	15 Lakhs	Industrial Espionage
23. Vendor Fraud	Criminal	Individual	Financial Loss	2023	Mumbai	Maharashtra	7 Lakhs	Social Engineering
24. Phishing Attack	Hacker	System Software	Loss of Data	2021	Chennai	TamilNadu	-	Malware
25. Online Identity Theft	Criminal	Individual	Financial Loss	2021	Pune	Pune	2.7 Lakhs	Identity theft
26. Business Email Scam	Criminal	Individual	Financial Loss	2022	New Delhi	Delhi	9 Lakhs	Social Engineering

Fig. 2 Overview of unstructured data

**Lemmatization:** It is done to normalize the text into root words for consistency within the dataset, enhancing the precision of further analyses [11].

$$x_T = \text{Tokenize}(x_i)$$

$$x_L = \text{StopWord}(x_T)$$

$$x_{Le} = \text{Lemmatize}(x_L)$$

$$D_{\text{preprocessed}} = (x_{1Le}, y_{1Le}), (x_{2Le}, y_{2Le}), \dots (x_{nLe}, y_{nLe})$$

These preprocessing techniques are at the front line of data standardization to a state that is effective for analysis; hence, the machine learning algorithm will interpret and classify instances of cybercrime with speed. Preprocessing provides a baseline for strong and reliable insights into the classification of cybercrime through tokenization and lemmatization. At the preprocessing stage, the processed data is split into two major portions: training data and testing data.

$$D = D_{\text{train}}, D_{\text{test}}$$

### 3.3 Feature Extraction

The key process that occurs in this phase is acquiring relevant features for cybercrime-related classification from the preprocessed data. This is accomplished using a process known as the Term Frequency-Inverse Document Frequency (TF-IDF) transformation [12, 13]. TF-IDF quantifies the importance prior connected to a given word in a document relative to a collection of documents. It considers the frequency of a word appearing in a document (TF) in addition to how rare the word appears in the data collection (IDF). This transformation identifies relevant and important keywords and phrases related to specific cyber threats, while increasing the potential of the feature set being discriminative.

$$\text{TF-IDF}_{ij} = \text{TF}_{ij} * \text{IDF}_i \quad (1)$$

Term Frequency (TF) represents the frequency of a term  $x_i$  in a document  $d_j$

$$\text{TF}_{ij} = \frac{x_i}{\text{Total words in } d_j} \quad (2)$$

where  $n_i$  is the number of times the word  $w_i$  in a document  $d_j$ .

Inverse Document Frequency (IDF) measures how rare a term  $w_i$  is across all documents in the collection

$$\text{IDF}_i = \log \frac{|D|}{|d : w_i \in d|} \quad (3)$$

where  $|D|$  is the total number of documents in the collection, and  $|d : w_i \in d|$  is the number of documents containing term  $w_i$ .

### 3.4 Model Training

During the model training stage, machine learning models such as SVM, Random Forests, XGBoost, and Naive Bayes will get trained on the preprocessed dataset [12, 13]. More specifically, we will take labeled examples of instances of cybercrime and feed it into the algorithms to learn relationships and patterns from the dataset, while also optimizing models' parameters in the process. Therefore, cross-validation and optimizations will be used to help tune the models and minimize or prevent overfitting the algorithms. After training, the model will be evaluated and tested in the next stage. The training process is not a single pass over the labeled a dataset; in fact, we will iteratively train the models over the same dataset to fine-tune the parameters and improve performance in rigorously classifying cybercrimes.

### 3.5 Model Testing

In the testing phase that comes after the training phase, an analysis is made on the performance of the trained models for obvious reasons to determine their suitability for real-world applications. The models' abilities and generalization against different cybercriminal activities are determined through testing the models on various realistic problems. By deploying the measurement concepts that include accuracy, precision, recall, and  $F1$  score, it becomes possible to quantify the ability of the model in performing the classification, while qualitative analysis examines the patterns and reasoning in arriving at the decisions and or classifications. This stage entails the testing of the models, and the knowledge gained is used to improve the models making it easier to counter digit security threats and strengthening security systems.

Evaluate the performance of each model using testing data ( $D_{test}$ ).

$$P_{alg} = \text{Classify(alg, } D_{test}) \text{ for alg } \in \{\text{RFC, SVM, NB, XG}\} \quad (4)$$

### 3.6 Analysis

The result analysis phase involves scrutiny of the model testing performance metrics to gain insights and make conclusions. Comparing and contrasting are used to assess the performance of various machine learning techniques in scenario categorization for cybercrimes. The peculiarities of each algorithm are discussed, and guidelines for choosing the most effective models for deployment in practical applications are provided. Furthermore, the implications of the study are highlighted while the performance and the classification of the results are presented in the results and performance analysis section.

## 4 Results

The results section serves as the nexus of the study, unveiling the classification outcomes of cybercrimes to the provided dataset. Figures 3 and 4 serve as visual narratives, encapsulating the classification results with a precision that illuminates the intricacies of cybercrime classification. The classification format employed by SVM, RFC, and Naive Bayes adheres to a standardized framework, as elucidated in Fig. 3. Conversely, the XGBoost algorithm assigns a specific cluster number to each cybercrime, indicative of its classification within the XGBoost framework, a method aptly showcased in Fig. 4. The misclassified results, showcased in Fig. 5, offer critical insights into the limitations and challenges faced by the machine learning algorithms in accurately classifying cybercrimes. By highlighting instances where the algorithms failed to correctly identify and classify cyber threats, these findings underscore the need for ongoing refinement and optimization of classification models.

The graph illustrating year-wise cyber-attacks is described in Fig. 6, providing a visual depiction of temporal trends in cybercrime incidents over time. Notably, the data reveals that the year 2022 has witnessed a notable increase in cybercrime

Year	State	Attack	Predicted Attack
2022	Delhi	Social Engineering	Social Engineering
2023	Karnataka	Identity theft	Identity theft
2022	UP	Social Engineering	Social Engineering
2020	Karnataka	Social Engineering	Social Engineering
2022	Andhra Pradesh	Social Engineering	Social Engineering
...	...	...	...
2021	Pune	Social Engineering	Social Engineering
2021	Maharashtra	Identity theft	Identity theft
2022	Haryana	Identity theft	Identity theft
2022	Goa	Social Engineering	Social Engineering
2022	West Bengal	Social Engineering	Social Engineering

**Fig. 3** Classification of cybercrimes using Support Vector Machine (SVM)

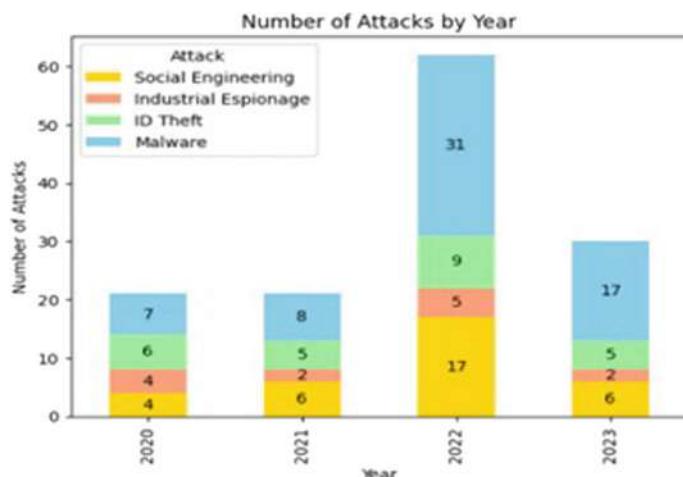
Year	State	Attack	Predicted Attack
2022	Delhi	3	3
2023	Karnataka	0	0
2022	UP	3	3
2020	Karnataka	3	3
2022	Andhra Pradesh	3	3
...	...	...	...
2021	Pune	3	3
2021	Maharashtra	0	0
2022	Haryana	0	0
2022	Goa	3	3
2022	West Bengal	3	3

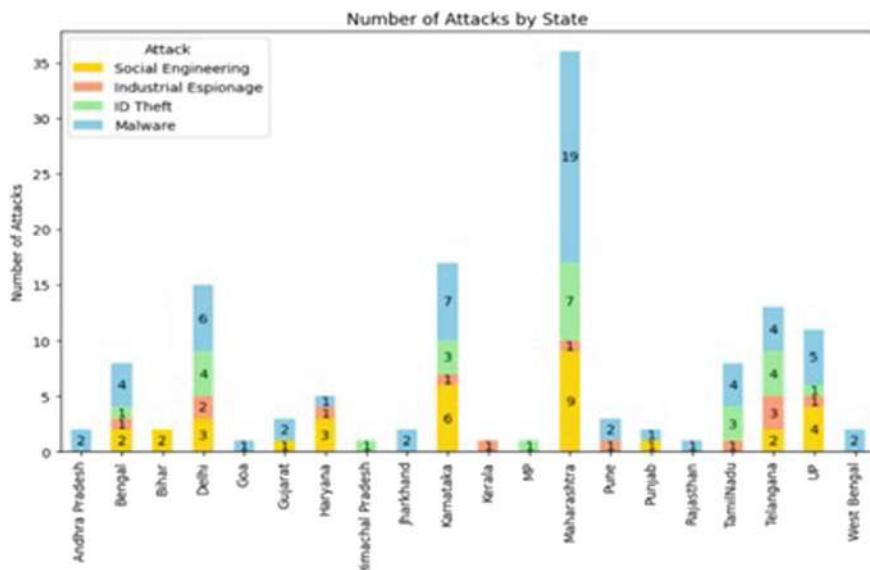
**Fig. 4** Classification of cybercrimes using XGBoost

Attack	Predicted Attack	Attack	Predicted Attack
403	1	2	
126	1	3	
422	2	1	
377	3	1	403 Industrial Espionage Malware
275	2	3	126 Industrial Espionage Social Engineering
286	0	3	422 Malware Industrial Espionage
437	3	1	286 Identity theft Social Engineering
383	1	2	383 Industrial Espionage Malware

(a)

(b)

**Fig. 5** Misclassified cybercrimes using **a** XGBoost, **b** SVM**Fig. 6** Year-wise stats of cyber crimes



**Fig. 7** State-wise stats of cyber crimes

incidents compared to previous years. This observation underscores the upgrading nature of cyber threats and the pressing need for robust cybersecurity measures to mitigate risks and safeguard digital assets.

The graph illustrating state-wide cyber-attacks is described in Fig. 7, shedding light on the distribution of cybercrime incidents across various regions. Notably, Maharashtra emerges as the top contributor to cybercrime incidents, exhibiting a significant concentration of cyber-attacks compared to other states. This observation underscores the importance of targeted cybersecurity measures in Maharashtra to address vulnerabilities and mitigate risks effectively.

## 5 Performance Analysis

The performance analysis section serves as a crucial component of the study, offering a meticulous evaluation of the efficiency of machine learning algorithms in cyber-crime classification. Presented both in a tabular and narrative format, quantitative and qualitative findings of classification performances, statistical measures, and evaluation criteria enhance the understanding of benefits and drawbacks of each algorithm. In the performance analysis section of the paper, several evaluation measures are used to evaluate the performance of developed ML models for the classification of cybercrimes. According to computing measures, these algorithms predict quantita-

tive measures for classification accuracy, precision score, recall score, and F1-score which are important to measure the performance of algorithms.

Accuracy is a metric which represents the ratio of correctly predicted instances to the total number of instances evaluated.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{False Positive} + \text{True Negative} + \text{False Negative}} \quad (5)$$

Precision measures the accuracy of positive predictions made by the model. It is the ratio of true positives to all instances predicted as positive.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (6)$$

Recall measures the ability of the model to correctly identify positive instances. It is the ratio of true positives to all actual positive instances in the dataset.

$$\text{Recall} = \frac{\text{TP}}{\text{True Positive} + \text{False Negative}} \quad (7)$$

F1-score is the harmonic mean of precision and recall, and it quantifies a model's accuracy in binary classification tasks. To provide a single number that sums up the model's performance, it combines the two metrics.

$$\text{F1 - Score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (8)$$

In Fig. 8, graphical displays of the number representations of the various algorithms are displayed to show their performance. However, Fig. 9 provides a bar chart of the performance metrics, which gives an insight on how well each algorithm performs as compared to others.

The confusion matrices in Fig. 10 have given an elaborate representation of the classification results beyond just the features of accuracy. When dealing with such specific values as TP, FP, TN, FN, stakeholders are able to find intricate patterns in the function of the algorithm.

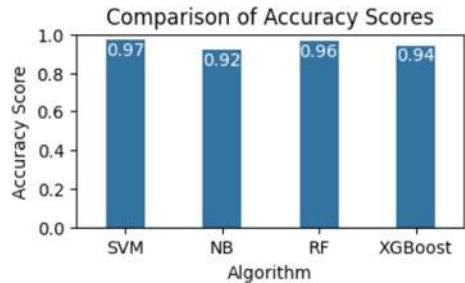
## 6 Conclusion

In conclusion, the performance analysis section sums up the whole investigation about the efficiency of the proposed ML algorithm in the cybercrime classification process. The information disseminated based on such analysis provides the necessity for the constant improvement of the cybersecurity environment, enabling the interested parties to prevent emerging threats and protect digital environments as effectively as possible. Firstly, there is an opportunity to enhance the CNNs and

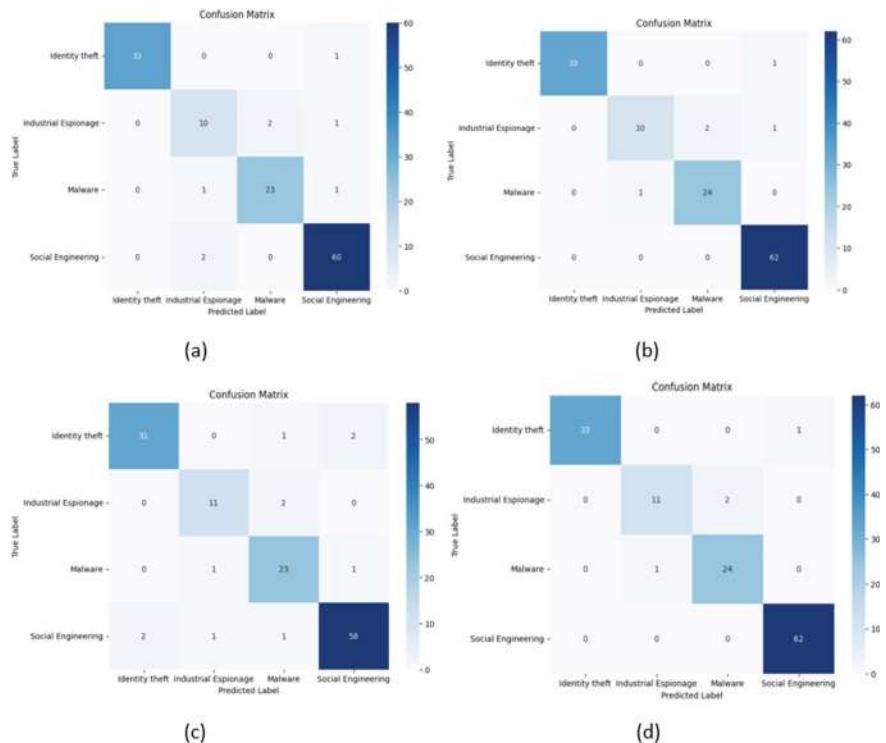
Naive Bayes Accuracy: 0.92	SVM Accuracy: 0.97
Naive Bayes Classification Report:	SVM Classification Report:
precision    recall    f1-score    support	precision    recall    f1-score    support
Identity theft    0.94    0.91    0.93    34	Identity theft    1.00    0.97    0.99    34
Industrial Espionage    0.85    0.85    0.85    13	Industrial Espionage    0.92    0.85    0.88    13
Malware    0.85    0.92    0.88    25	Malware    0.92    0.96    0.94    25
Social Engineering    0.95    0.94    0.94    62	Social Engineering    0.98    1.00    0.99    62
accuracy	accuracy
macro avg    0.90    0.90    0.90    134	macro avg    0.96    0.94    0.95    134
weighted avg    0.92    0.92    0.92    134	weighted avg    0.97    0.97    0.97    134
Random Forest Accuracy: 0.96	XGBoost Accuracy: 0.94
Random Forest Classification Report:	XGBoost Classification Report:
precision    recall    f1-score    support	precision    recall    f1-score    support
Identity theft    1.00    0.97    0.99    34	0    1.00    0.97    0.99    34
Industrial Espionage    0.91    0.77    0.83    13	1    0.77    0.77    0.77    13
Malware    0.92    0.96    0.94    25	2    0.92    0.92    0.92    25
Social Engineering    0.97    1.00    0.98    62	3    0.95    0.97    0.96    62
accuracy	accuracy
macro avg    0.95    0.92    0.94    134	macro avg    0.91    0.91    0.91    134
weighted avg    0.96    0.96    0.96    134	weighted avg    0.94    0.94    0.94    134

**Fig. 8** Performance metrics of proposed machine learning models

**Fig. 9** Accuracy comparison graph of proposed models



RNNs to increase the efficiency of the extraction of complex shapes and details of a diverse set of cybercrime indicators. Moreover, the inclusion of real-time data streams and the incorporation of dynamic feature extraction techniques would help the system react to changing threats as soon as possible. In summary, the concept proposed for halting cybercriminals' advancements through the implementation of a novel mechanism for classifying cybercrimes generates immense potential for the improvement of cybersecurity and the protection of digital environments against threat actors.



**Fig. 10** Confusion matrices of algorithms **a** XGBoost, **b** RFC, **c** Naive Bayes, **d** SVM

## References

- Horan C, Saiedian H (2021) Cyber crime investigation: landscape, challenges, and future research directions. *J Cybersecur Privacy* 1(4):580–596
- Phillips K, Davidson JC, Farr RR et al (2022) Conceptualizing cybercrime: definitions, typologies and taxonomies. *Forens. Sci.* 2(2):379–398
- Damaševičius R, Venčkauskas A, Toldinas J et al (2021) Ensemble-based classification using neural networks and machine learning models for windows PE malware detection. *Electronics* 10:485
- Djenna A, Barka E, Benchikh A et al (2023) Unmasking cybercrime with artificial-intelligence-driven cybersecurity analytics. *Sensors* 23:6302
- Asam M, Hussain SJ, Mohatram M et al (2021) Detection of exceptional malware variants using deep boosted feature spaces and machine learning. *Appl. Sci.* 11:10464
- Rashid MM, Kamruzzaman J, Hassan MM et al (2020) Cyberattacks detection in IoT-based smart city applications using machine learning techniques. *Int. J. Environ. Res. Public Health* 17:9347
- Gautam AK, Bansal A (2022) Performance analysis of supervised machine learning techniques for cyberstalking detection in social media. *J Theor Appl Inform Technol* 100(2):449–461
- Gutiérrez-Esparza GO, Vallejo-Allende M, Hernández-Torruco J (2019) Classification of cyber-aggression cases applying machine learning. *Appl. Sci.* 9:1828

9. Kumara BA, Kodabagi MM, Choudhury T et al (2021) Improved Email classification through an enhanced data preprocessing approach. *Spat. Inf. Res.* 29:247–255. <https://doi.org/10.1007/s41324-020-00378-y>
10. Rupa Ch, Gadekallu TR, Abidi MH, Al-Ahmari (2020) A computational system to classify cyber crime offenses using machine learning. *Sustainability* 12, 4087. <https://doi.org/10.3390/su12104087>
11. Nafis NSM, Awang S (2020) The impact of pre-processing and feature selection on text classification. In: Zakaria Z, Ahmad R (eds) *Advances in electronics engineering. Lecture notes in electrical engineering*, vol 619. Springer, Singapore. [https://doi.org/10.1007/978-981-15-1289-6\\_25](https://doi.org/10.1007/978-981-15-1289-6_25)
12. Lai C-M, Chen M-H, Kristiani E et al (2022) Fake news classification based on content level features. *Appl Sci* 12:1116
13. Sai Varshitha G, Rupa, Ch, Divya D (2024) Real time blood bank communication integrating access control using XGBoost and Supabase. In: 2024 Second international conference on data science and information system (ICDSIS), Hassan, India, pp 1–6. <https://doi.org/10.1109/ICDSIS61070.2024.10594708>

# Prognostic Analysis of Logistics Operation by Modeling Ship Berthing Problem



Phuoc Quy Phong Nguyen, Hoang Phuong Nguyen, Van Phuc Nguyen,  
Duc Chuan Nguyen, and Dang Khoa Pham Nguyen

**Abstract** Global trade relies on the efficiency of port operations and logistics; prompt vessel berthing is, therefore, necessary to lower delays and related expenses. Berthing timings can be influenced by elements including ship size berth capacity. Thus, improving port operations and resource allocation depends on precisely forecasting turnaround and waiting times. This paper addresses the creation of two machine learning models such as turnaround time and waiting time models by using Huber regression, Tweedie regression, and Gradient Boosted Regression (GBR) to anticipate operational metrics in port and logistics. The datasets for both models were split 80% for training and 20% for testing therefore guaranteeing strong evaluation. Although Tweedie regression presents a flexible approach by allowing different distributions including those with non-constant variance the Huber regression model shows strong resistance to outliers, hence controlling the variations in turnaround times. GBR uses ensemble learning to aggregate several weak learners, hence raising predicted accuracy. Regarding generalization and accuracy, performance tests show that the Tweedie and GBR models exceeded the Huber model. Particularly the Tweedie model obtained a training  $R^2$  of 0.9999 and a test  $R^2$  of 0.9671; GBR demonstrated a training  $R^2$  of 1.0. This paper shows how various machine learning methods might efficiently maximize port operations and offers insightful analysis for improving operational efficiency.

---

P. Q. P. Nguyen  
Faculty of Maritime, Maritime College II, Ho Chi Minh City, Vietnam  
e-mail: [phong@ut.edu.vn](mailto:phong@ut.edu.vn)

H. P. Nguyen  
Academy of Politics Region II, Ho Chi Minh City, Vietnam

V. P. Nguyen · D. C. Nguyen · D. K. P. Nguyen (✉)  
Institute of Maritime, Ho Chi Minh City University of Transport, Ho Chi Minh City, Vietnam  
e-mail: [khoapnd@ut.edu.vn](mailto:khoapnd@ut.edu.vn)

D. C. Nguyen  
e-mail: [chuan.nguyen@ut.edu.vn](mailto:chuan.nguyen@ut.edu.vn)

**Keywords** Machine learning · Port operation · Logistics management · Maritime transportation · Prediction

## 1 Introduction

Ports are critical infrastructures in global trade, acting as hubs for the movement of goods via maritime transportation [1, 2]. Globalization, the fast development of economies, and the growing reliance on containerized goods have all driven a tremendous need for port services [3, 4]. However, the increased marine shipping activities are also resulting in adverse environmental effects especially emissions from ship berthing [5, 6]. In the case of cargo handling, during ship refueling, or maintenance, ship berthing is the period a vessel spends parked in a port [7]. Ships need auxiliary engines to keep running the vital systems throughout this period. However, it leads to the emission of dangerous pollutants including carbon dioxide ( $\text{CO}_2$ ), nitrogen oxides ( $\text{NO}_x$ ), sulfur oxides ( $\text{SO}_x$ ), and particulate matter (PM). Thus, a major concern for environmental regulators and port authorities is the efficient berthing of cargo ships to avoid excessive emissions at this stage [8]. It is also worth noting that ship emissions have garnered a lot of attention but an increased intervention in ship activities is needed as it generates a large volume of greenhouse gases (GHGs) [9–11]. Auxiliary engines at port greatly affect the emissions of a ship at berth. When multiplied by the number of ships that call at major ports daily, this prolonged immobility period has a major environmental effect [8, 12]. Auxiliary engines run on fuel oil during berthing to provide necessary services including power generation, heating, and cooling as well as to help cargo handling activities. Dependency on these auxiliary engines can lead to notable fuel consumption, which directly affects ship running expenses. For ship operators, rising fuel use at berth can be a major cost driver given changing worldwide fuel prices, therefore affecting total running expenses [13, 14]. Economically speaking, ship operators are always trying to cut running expenses to keep profitability in a very competitive sector [15]. Particularly for vessels that spend more time at port due to congestion, long cargo handling procedures, or port service delays, fuel usage during the berthing process might be a neglected cost element [16, 17].

Because of the many factors involved like ship size, draft, depth and length of the berth, and kind of cargo, modeling ship berthing procedures is intrinsically difficult. These components interact in ways that can significantly influence operational outcomes including fuel use, waiting time, and turnaround time. Additional complicating the modeling approach are outside factors including port congestion, temperature, and port service effectiveness [18, 19]. It is observed that conventional modeling techniques often are unable to sufficiently capture the dynamic character of ship berthing procedures [20]. A deterministic model could, for example, ignore the variation in waiting times resulting from unstable handling of cargo or port congestion. Furthermore, challenging traditional models to fairly forecast operational results are the non-linear interactions among the many variables, such as the ship size and the

berth depth [21, 22]. Hence, the need to optimize environmental as well as financial goals adds another level of complication for cargo ship berthing problems. The environmental and economic challenges associated with ship berthing operations are significant, but ML offers a promising solution for modeling and optimizing these processes. ML can increase the accuracy of forecasts concerning fuel consumption, emissions, and operational efficiency by using previous data and sophisticated algorithms. Adopting ML approaches will be essential to guarantee sustainable and effective port operations as ports continue to face increasing demands and tighter environmental rules [23]. While reinforcement learning has great possibility for real-time optimization, supervised learning with its prediction powers is especially suited for modeling ship berthing activities. The literature review shows that ship berthing is a complex issue to manage and model. Hence, the present study is an endeavor to explore the ability of modern ML approaches like Huber, Gradient Boosting Regression (GBR), and Tweedie for the efficient prognostics of logistic ship berthing problems.

## 2 Materials and Methods

### 2.1 Data Collection and Analysis

Appropriately modeling and optimizing ship berthing processes depend on collected data for this purpose. In this instance, data on ship dimensions (e.g., length, draft), berth features (e.g., length, depth), and operational metrics (e.g., turnaround time, waiting time) came from records of port activity. Accurate machine learning models require these predictors and response variables. The data is collected from manual logs of ship operations, automated tracking systems (like AIS), and port management systems. Reliable models depend on high-quality data. To guarantee consistency, the gathered data went through preprocessing processes including feature scaling, outlier detection, and management of missing values. Better knowledge of operational trends made possible by proper data analysis was essential for training machine learning algorithms able to predict and maximize berthing procedures.

### 2.2 ML Approaches

#### 2.2.1 Huber Regression

Huber regression makes use of least-squares and median regression to perform as an efficient ML approach. Huber is resilient to data outliers. Since the extreme outliers can disproportionately affect the model to produce poor predictions in the case of

conventional linear regression, the Huber regression introduces a quadratic loss function for small errors and a linear loss function for significant errors. In this way, it takes care of both these issues. This approach ensures that, unlike in standard regression, small errors are handled as such and that outliers affect the results only marginally [24, 25]. Huber regression aids in ship berthing operations where data may include severe outliers resulting from unusual port delays or operational inconsistencies. Employing the removal of outliers, the model provides more reliable forecasts for typical port operations, therefore enabling more consistent port management decisions and fuel usage estimates optimization.

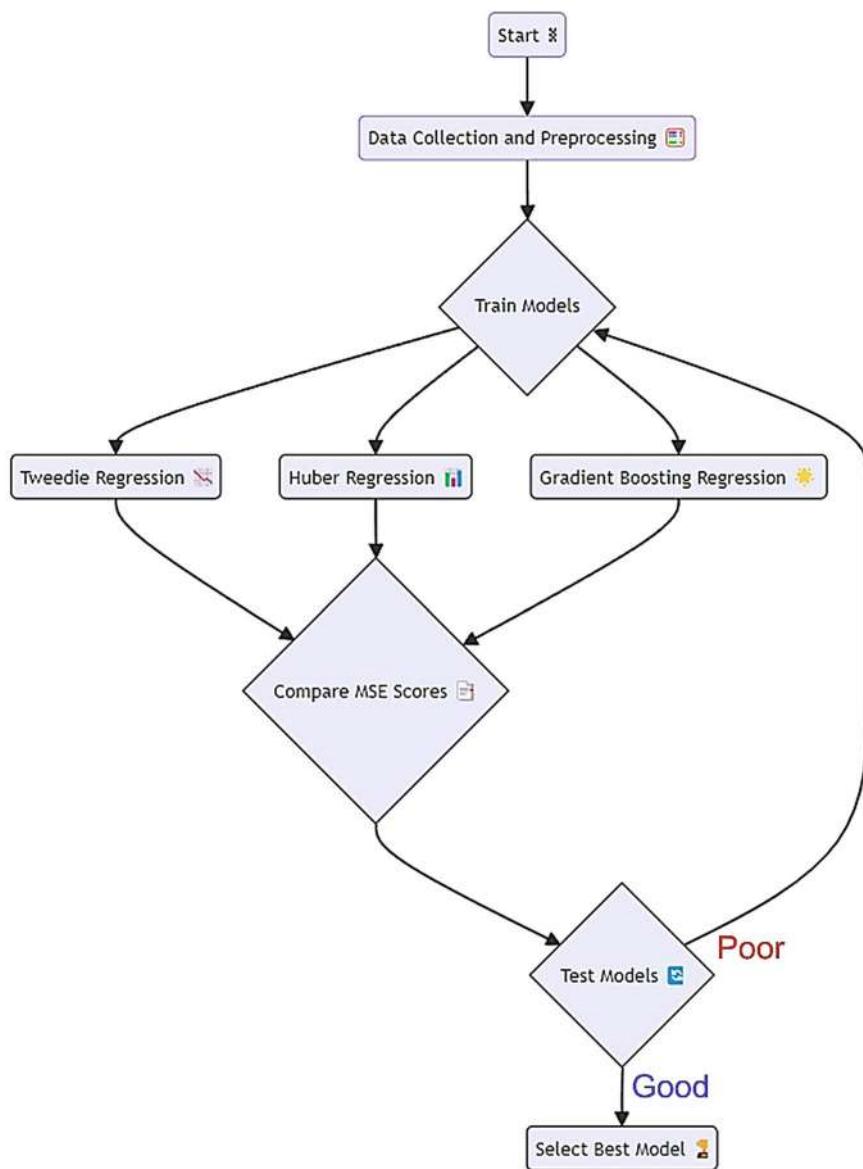
### 2.2.2 Gradient Boosting Regression

Gradient boosting regression is a powerful ensemble learning method. It is used to build predictive models by sequentially training weak learners. These weak learners are typically decision trees. The fundamental concept is to iteratively modify the prior models in the direction of the gradient of the loss function thereby reducing their residual errors [26, 27]. As the model emphasizes fixing larger errors, it becomes more accurate over time. Gradient boosting regression can capture complicated non-linear interactions in ship berthing operations, taking into account parameters such as ship size, berth depth, and operational delays. When several predictors reveal interactions, the approach is very helpful for handling heterogeneous data [28, 29]. It is a good choice for projecting environmental consequences and operational performance at ports since it can manage lacking data and generate excellent predictions with great accuracy. Particularly in real world, complicated situations like port management, gradient boosting is quite adaptable and usually produces outstanding predictive performance.

### 2.2.3 Tweedie Regression

Tweedie regression is a specialized statistical technique used to model data that has a combination of discrete and continuous components, often observed in domains like insurance claims, environmental science, and energy consumption. Particularly appropriate for managing data with skewed distributions, such as whereby responses can be zero or positive (e.g., ship berthing delays or fuel usage with occasional zero values), the Tweedie distribution belongs to the exponential family [30, 31]. Tweedie regression gives flexibility in machine learning by letting the user modify the “power” value, hence regulating the variance-to-mean relationship of the model. This makes it perfect for simulating complicated situations in port operations, where non-normal behavior of pollutants, fuel consumption, or waiting times may arise. By better capturing the quirks of ship operational data, Tweedie regression can raise prediction accuracy in port management models. For jobs like estimating fuel consumption and pollution outputs during berthing operations, its capacity to manage mixed data

distributions makes it especially valuable. The flow chart for ML implementation in this study is depicted in Fig. 1.



**Fig. 1** Schematics of ML implementation

## 2.3 Evaluation Criteria

The assessment of machine learning models depends on their accuracy and dependability, so their evaluation is essential. Typical measures are coefficient of determination ( $R^2$ ), mean squared error (MSE), and mean absolute percentage error (MAPE).  $R^2$  gauges how much the independent factors explain of variance in the dependent variable. Its values lie between 0 and 1; values nearer 1 indicate improved model performance [32]. A  $R^2$  of 0 indicates that the model fails to explain any data variability. MSE computes the average squared difference between actual and expected data, therefore illuminating the error scale of the model. Because errors are penalized quadratically, lower MSE values indicate improved model performance. Conversely, MAPE gauges the average absolute percentage change between actual and expected values. When assessing forecasting models, this statistic is especially helpful since it offers inaccuracy in percentage terms, which facilitates the understanding of several datasets.

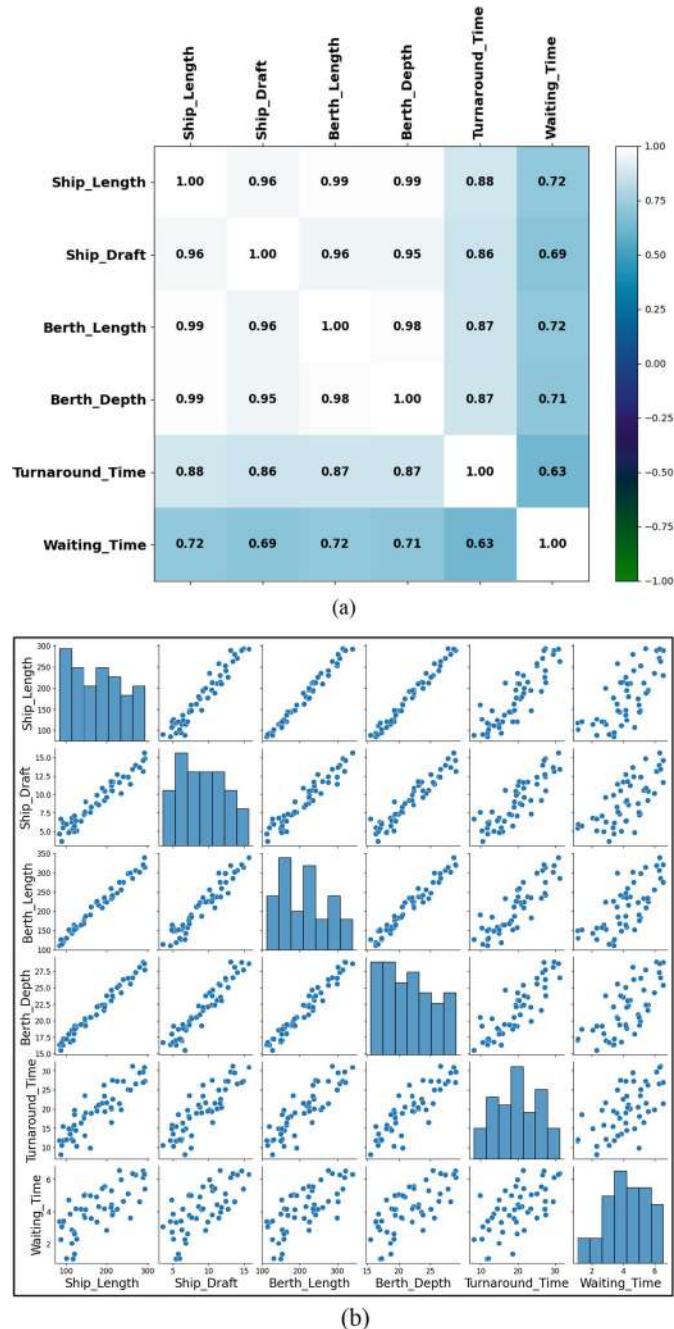
## 3 Results and Discussion

### 3.1 Data Analysis and Correlation

The given Table 1 and Fig. 2a offer an understanding of the relationships among the several variables by displaying their correlation. From  $-1$  to  $1$ , correlation values run; values nearer  $1$  point to a strong positive link. For example, the  $0.96$  connection between ship length and ship draft indicates that, given more underwater clearance, larger ships typically have deeper drafts. Larger ships are usually accommodated at longer berths, which is operationally sensible indicated by an exceptionally high correlation of  $0.99$  between ship length and berth length. Larger ships thus need deeper berths for safe docking, as the significant correlation of  $0.99$  between ship length and berth depth indicates. Longer ships presumably take more time for loading and unloading operations based on the positive correlation of

**Table 1** Correlation matrix

	Ship length	Ship draft	Berth length	Berth depth	Turnaround time	Waiting time
Ship length	1	0.96	0.99	0.99	0.88	0.72
Ship draft	0.96	1	0.96	0.95	0.86	0.69
Berth length	0.99	0.96	1	0.98	0.87	0.72
Berth depth	0.99	0.95	0.98	1	0.87	0.71
Turnaround time	0.88	0.86	0.87	0.87	1	0.63
Waiting time	0.72	0.69	0.72	0.71	0.63	1



**Fig. 2** Correlation **a** heatmap **b** pair plot

0.88 between ship length and turnaround time. By comparison, at 0.72 the association between ship length and waiting time is modest, suggesting that bigger ships may have longer waiting times, maybe because of restricted availability of adequately big berths. With a connection of 0.63 between turnaround time and waiting time, greater turnaround times seem to somewhat raise waiting times depending on different operational conditions. The table shows generally that ship size significantly influences port infrastructure; operational durations have only a modest impact, therefore underlining the complexity of port logistics and the interaction between ship sizes and port operations. These relationships highlight the requirement for effective design and planning in port facilities to allow bigger boats and maximize turnaround and waiting periods for maximum operational effectiveness. The data distribution and trends are depicted in Fig. 2b as a correlation pair plot.

### 3.2 Turnaround Time Model

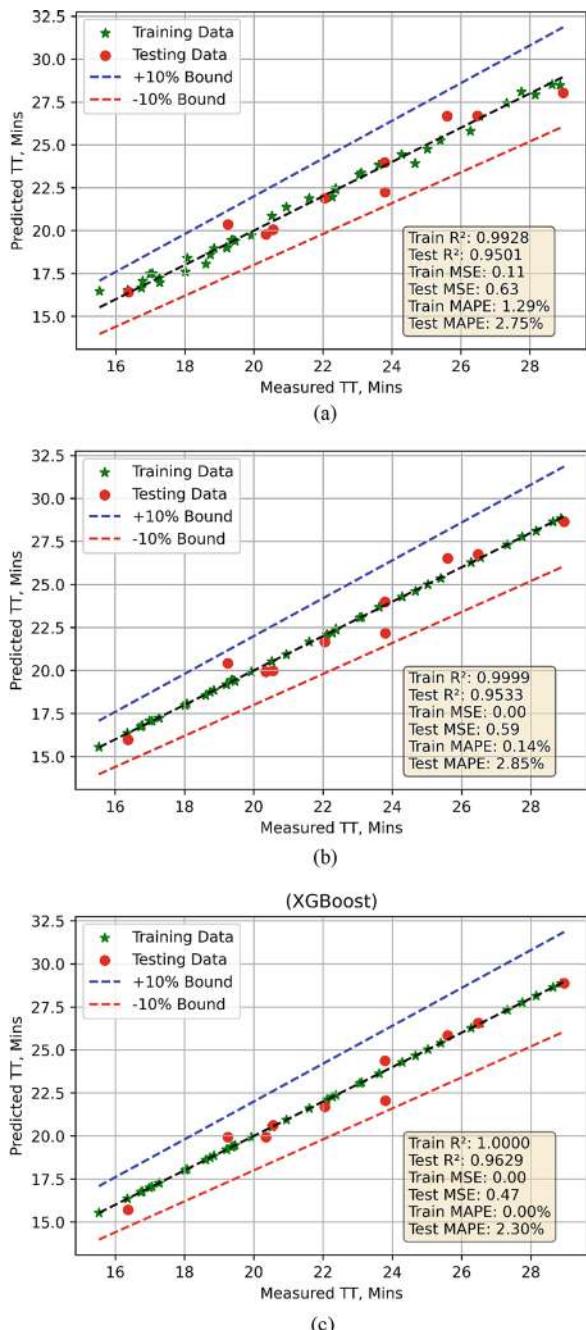
Using several machine learning techniques more especially, Huber regression, Tweedie regression, and GBR the turnaround time model precisely forecasts operational turnaround times. To guarantee strong evaluation, the dataset was split into an 80% training set and a 20% testing set. Combining aspects of linear regression and resilient loss functions helps Huber regression known for its resilience to outliers effectively manages the variability in turnaround times. Because it fits many distributions, including those with non-constant variance, which lets a more flexible model fit the kind of data, Tweedie regression is very helpful for modeling turnaround times. By aggregating the strengths of several weak learners, GBR using ensemble learning increases predicted accuracy, hence improving performance in capturing the intricate relationships within the dataset. These approaches taken together enable a thorough investigation, which finally produces a consistent model for turning around times that can be applied to maximize port operations and improve general efficiency.

Turning around times are effectively predicted by the Huber, Tweedie, and Gradient Boosted Regression (GBR) models as model performance measures are listed in Table 2. With a far higher test MSE of 0.6320, the Huber model's training mean squared error (MSE) of 0.1053 points to a possible overfitting problem. The model's comparative performance is illustrated in Fig. 3a. Notwithstanding this, the model demonstrates great predictive performance with a training  $R^2$  of 0.9928 and a test  $R^2$  of 0.9501, therefore explaining a great share of variation in the training set.

**Table 2** Model evaluation for turnaround time

Model	Train MSE	Test MSE	Train $R^2$	Test $R^2$	Train MAPE, %	Test MAPE, %
Huber	0.1053	0.6320	0.9928	0.9501	1.2920	2.74
Tweedie	0.0015	0.5919	0.9999	0.9533	0.1439	2.85
GBR	0	0.4703	1	0.9629	0.0049	2.3

**Fig. 3** Turnaround time model's comparative performance for **a** Huber  
**b** Tweedie **c** GBR



With a training MSE of 0.0015 and a test MSE of 0.5919, the Tweedie model (Fig. 3b) shows really good performance. While the test  $R^2$  of 0.9533 shows great generalizing to unseen data, its training  $R^2$  value of 0.9999 demonstrates an outstanding fit to the training data. With a test MAPE of 2.85%, the Train MAPE of 0.1439 reveals just a minimum average percentage error. With a training MSE of 0 and a test MSE of 0.4703, the GBR model fits the training data. With a test MAPE of 2.3%, and its training  $R^2$  of 1.0 the GBR-based model (Fig. 3c) shows excellent prediction on training data; its test  $R^2$  of 0.9629 shows strong performance on test data, so stressing its accuracy. The Tweedie and GBR models show generally better performance, especially in terms of generalization and precision in turning around times.

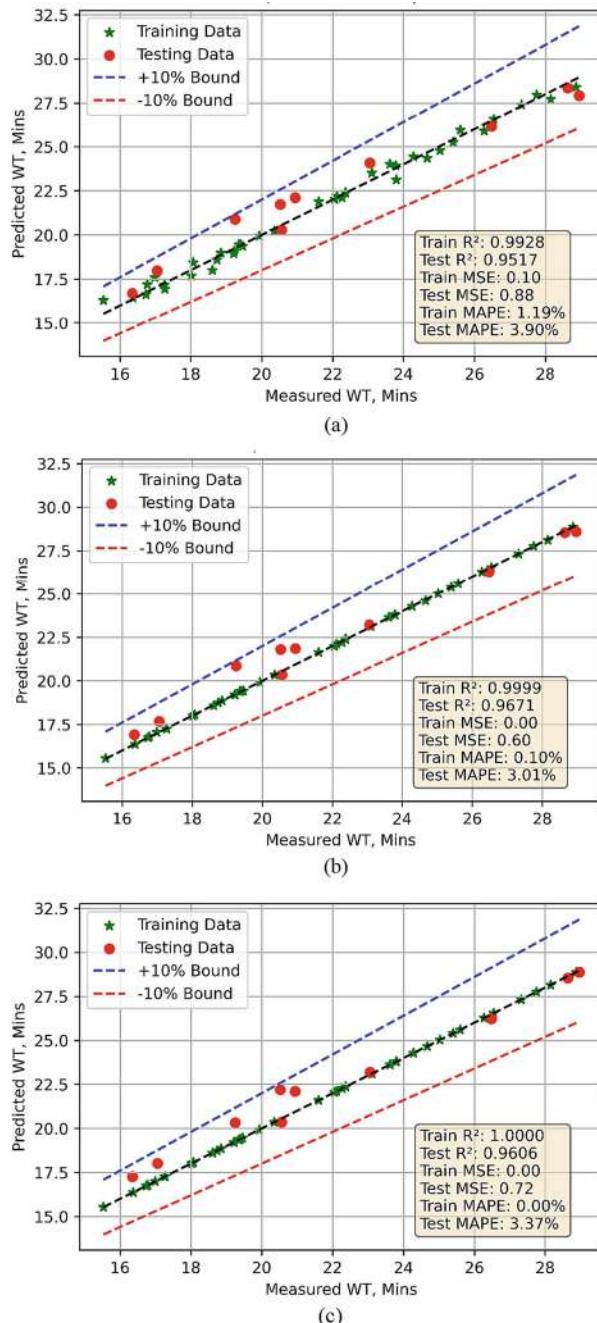
### 3.3 Waiting Time Model

The waiting time model efficiently estimates operational waiting times employing several machine learning methods, namely Huber regression, Tweedie regression, and Gradient Boosted Regression (GBR). Since Tweedie regression allows several distributions, including those with non-constant variance, therefore enables a more flexible approach to fit the data features. It is therefore quite helpful for modeling waiting times. GBR uses ensemble learning concurrently to improve prediction accuracy by combining the strengths of numerous weak learners, hence enhancing its effectiveness in capturing the complex interactions in the dataset. These approaches taken together allow a thorough study, which produces a dependable model for estimating waiting times that can maximize port operations and improve general efficiency. Table 3 summarizes how well the Huber, Tweedie, and GBR models predict waiting times.

Considering the training performance is much better than the testing performance, the Huber model attained a training MSE of 0.097 and a test MSE of 0.88, suggesting a minor overfitting problem, as depicted in Fig. 4a. With a training  $R^2$  of 0.9928 and a test  $R^2$  of 0.9517, the model does, however, retain significant predictive ability and shows that it explains a good amount of the variance in the training data. Reflecting good accuracy, the test MAPE is 3.89%; the training MAPE is 1.19%. With a training MSE of 0.0008 and a test MSE of 0.60, the Tweedie model in Fig. 4b, shows robust performance. While the test  $R^2$  of 0.9671 shows remarkable generalizing to new, unknown data, its training  $R^2$  value of 0.9999 demonstrates a near-perfect fit to the

**Table 3** Model evaluation for waiting time

Model	Train MSE	Test MSE	Train $R^2$	Test $R^2$	Train MAPE, %	Test MAPE, %
Huber	0.097	0.88	0.9928	0.9517	1.19	3.89
Tweedie	0.0008	0.60	0.9999	0.9671	0.1	3.01
GBR	0	0.72	1	0.9606	0.0042	3.37



**Fig. 4** Waiting time model's comparative performance for **a** Huber **b** Tweedie **c** GBR

training data. Further verifying its dependability and precision are the Train MAPE of 0.1% and the test MAPE of 3.01%. With a training MSE of 0 and an  $R^2$  of 1.0, the GBR model as depicted in Fig. 4c shows excellent training performance and produces perfect forecasts on the training set. Strong performance on the test is rather accurate in estimating waiting times. Generally speaking, the Tweedie and GBR models show better performance than the Huber model, especially concerning generalization and accuracy in estimating operational waiting times, providing useful instruments for optimizing port operations.

## 4 Conclusion

The study emphasizes the effectiveness of Huber regression, Tweedie regression, and Gradient boosting regression (GBR) models for estimating turnaround and waiting times. With a training  $R^2$  of 0.9928 and a test  $R^2$  of 0.9501, the Huber model showed a training MSE of 0.1053 and a test MSE of 0.6320, therefore suggesting a possible overfitting issue in the turnaround time model. This implies that although the model performs somewhat poorly on unknown data, it is rather good in training. With a stunning training MSE of 0.0015 a test MSE of 0.5919 and a training  $R^2$  of 0.9999 and a test  $R^2$  of 0.9533, the Tweedie model produced outstanding performance. Though its test MSE of 0.4703 and test  $R^2$  of 0.962 demonstrate its outstanding predictive abilities, the GBR model obtained faultless training performance with an MSE of 0 and a training  $R^2$  of 1.0. Comparably, in the waiting time model, the Huber model recorded a training MSE of 0.097 and a test MSE of 0.88, implying a moderate overfitting problem with a training  $R^2$  of 0.9928 and a test  $R^2$  of 0.9517. With a training MSE of 0.0008 and a test MSE of 0.60, the Tweedie model once more excelled showing a near-perfect match with a training  $R^2$  of 0.9999. Perfect predictions on training data were indicated by the GBR model's sustained 0 training MSE and 1.0 training  $R^2$ . Particularly in terms of generality and accuracy, the Tweedie and GBR models were superior to the Huber model, therefore proving their indispensable nature as fundamental instruments for improving operational efficiency within port management systems.

## References

1. Hu B (2018) Application of evaluation algorithm for port logistics park based on PCA-SVM model. Polish Marit Res 25:29–35. <https://doi.org/10.2478/pomr-2018-0109>
2. Nguyen HP, Nguyen PQP, Nguyen DKP, Bui VD, Nguyen DT (2023) Application of IoT technologies in seaport management. JOIV Int J Informatics Vis 7:228–40. <https://doi.org/10.30630/joiv.7.1.1697>
3. Gucma S (2019) Conditions of Safe ship operation in seaports—optimization of port waterway parameters. Polish Marit Res 26:22–9. <https://doi.org/10.2478/pomr-2019-0042>

4. Sharif MB, Gorbanpour AH, Ghassemi H, He G (2023) Investigating the harbour basin tranquillity in the genaveh port development plan. Polish Marit Res 30:145–55. <https://doi.org/10.2478/pomr-2023-0015>
5. Bojić F, Bošnjak R, Gudelj A (2021) Review of smart ports in the European union
6. Anastasopoulos AT, Sofowote UM, Hopke PK, Rouleau M, Shin T, Dheri A et al (2021) Air quality in Canadian port cities after regulation of low-sulphur marine fuel in the North American emissions control area. Sci Total Environ 791:147949. <https://doi.org/10.1016/j.scitotenv.2021.147949>
7. Hoang AT, Foley AM, Nižetić S, Huang Z, Ong HC, Ölcer AI et al (2022) Energy-related approach for reduction of CO<sub>2</sub> emissions: a critical strategy on the port-to-ship pathway. J Clean Prod 355:131772. <https://doi.org/10.1016/j.jclepro.2022.131772>
8. Villalba G, Gemechu ED (2011) Estimating GHG emissions of marine ports—the case of Barcelona. Energy Policy 39:1363–8
9. Livanou S, Chatzistelios G, Lyridis DV, Bellos E (2022) LNG vs. MDO in Marine Fuel Emissions Tracking. Sustainability 14:3860. <https://doi.org/10.3390/su14073860>
10. Pham VV, Hoang AT (2019) Technological perspective for reducing emissions from marine engines. Int J Adv Sci Eng Inf Technol 9:1989–2000. <https://doi.org/10.18517/ijaseit.9.6.10429>
11. Hoang AT, Pandey A, Martinez De Osés FJ, Chen W-H, Said Z, Ng KH, et al (2023) Technological solutions for boosting hydrogen role in decarbonization strategies and net-zero goals of world shipping: challenges and perspectives. Renew Sustain Energy Rev 188:113790. <https://doi.org/10.1016/j.rser.2023.113790>
12. Ben-Hakoun E, Van De Voorde E, Shifman Y (2022) Trends in emission inventory of marine traffic for port of Haifa. Sustainability 14:908. <https://doi.org/10.3390/su14020908>
13. Knežević V, Radonja R, Dundović Č (2018) Emission inventory of marine traffic for the port of Zadar. Pomorstvo 32:239–44. <https://doi.org/10.31217/p.32.2.9>
14. Stazić L, Radonja R, Pelić V, Lalić B (2020) The port of split international marine traffic emissions inventory. Pomorstvo 34:32–9. <https://doi.org/10.31217/p.34.1.4>
15. Rudzki K, Gomulka P, Hoang AT (2022) Optimization model to manage ship fuel consumption and navigation time. Polish Marit Res 29:141–53. <https://doi.org/10.2478/pomr-2022-0034>
16. Lee H-T, Lee J-S, Son W-J, Cho I-S (2020) Development of machine learning strategy for predicting the risk range of ship's berthing velocity. J Mar Sci Eng 8:376. <https://doi.org/10.3390/jmse8050376>
17. Senol YE, Seyhan A (2024) A novel machine-learning based prediction model for ship manoeuvring emissions by using bridge simulator. Ocean Eng 291:116411. <https://doi.org/10.1016/j.oceaneng.2023.116411>
18. Khan RU, Yin J, Mustafa FS, Shi W (2023) Factor assessment of hazardous cargo ship berthing accidents using an ordered logit regression model. Ocean Eng 284:115211. <https://doi.org/10.1016/j.oceaneng.2023.115211>
19. Moorthy R, Teo C-P (2006) Berth management in container terminal: the template design problem. OR Spectr 28:495–518. <https://doi.org/10.1007/s00291-006-0036-5>
20. Azhary S, Purwanto DB, Nurhadi H, Pramujati B, Effendi MK, Widjaja S, et al (2021) Design of remotely operated vehicle prototype for ship biofouling inspection on berth. In: 2021 International Conference Advance Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA). IEEE, pp 223–228. <https://doi.org/10.1109/ICAMIMIA54022.2021.9807777>
21. Nguyen V-S, Im N-K (2019) Automatic ship berthing based on fuzzy logic. Int J FUZZY Log Intell Syst 19:163–71. <https://doi.org/10.5391/IJFIS.2019.19.3.163>
22. Im N-K, Nguyen V-S (2018) Artificial neural network controller for automatic ship berthing using head-up coordinate system. Int J Nav Archit Ocean Eng 10:235–49. <https://doi.org/10.1016/j.ijnaoe.2017.08.003>
23. Filom S, Amiri AM, Razavi S (2022) Applications of machine learning methods in port operations—a systematic literature review. Transp Res Part E Logist Transp Rev 161:102722. <https://doi.org/10.1016/j.tre.2022.102722>
24. Sun Q, Zhou W-X, Fan J (2020) Adaptive huber regression. J Am Stat Assoc 115:254–65. <https://doi.org/10.1080/01621459.2018.1543124>

25. Feng Y, Wu Q (2022) A statistical learning assessment of Huber regression. *J Approx Theory* 273:105660. <https://doi.org/10.1016/j.jat.2021.105660>
26. Sibindi R, Mwangi RW, Waititu AG (2023) A boosting ensemble learning based hybrid light gradient boosting machine and extreme gradient boosting model for predicting house prices. *Eng Reports* 5. <https://doi.org/10.1002/eng2.12599>
27. Bentéjac C, Csörgő A, Martínez-Muñoz G (2021) A comparative analysis of gradient boosting algorithms. *Artif Intell Rev* 54:1937–67. <https://doi.org/10.1007/S10462-020-09896-5/METRICS>
28. Su Y (2020) Prediction of air quality based on gradient boosting machine method. In: 2020 International Conference Big Data Information Education, IEEE, pp 395–397. <https://doi.org/10.1109/ICBDIE50010.2020.00099>
29. Kefalas A, Ofner AB, Pirker G, Posch S, Geiger BC, Wimmer A (2022) Estimation of combustion parameters from engine vibrations based on discrete wavelet transform and gradient boosting. *Sensors* 22:4235. <https://doi.org/10.3390/s22114235>
30. Petterle RR, Bonat WH, Kokonendji CC, Seganfredo JC, Moraes A, da Silva MG (2019) Double poisson-tweedie regression models. *Int J Biostat* 15. <https://doi.org/10.1515/ijb-2018-0119>
31. Bonat WH, Kokonendji CC (2017) Flexible Tweedie regression models for continuous data. *J Stat Comput Simul* 87:2138–52. <https://doi.org/10.1080/00949655.2017.1318876>
32. Gopi A, Sharma P, Sudhakar K, Nguí WK, Kirpichnikova I, Cuce E (2022) Weather impact on solar farm performance: a comparative analysis of machine learning techniques. *Sustainability* 15:439. <https://doi.org/10.3390/su15010439>

# A Virtual Cognitive Intelligence Framework for Digital Telehealth Zone



Rohit Agarwal, Amit Kumar Sinha, Utpala Dutta, Lu Wang,  
and Bharati Rathore

**Abstract** Cognitive computing technologies are gradually becoming integrated with remote healthcare systems in the last years, exhibiting promising capabilities to revolutionize patient care delivery even in remote or underserved areas. In this paper, a wide study of the literature in which cognitive computing is applied to remote healthcare, so covering a few domains like diagnostic support, remote monitoring, personalized treatment planning, virtual assistants, health data analysis, natural language processing, decision support systems, medical imaging analysis, and remote education and training is presented. Based upon a systematic review of the existing research findings, this paper summarizes the current technological superiority, identifying the pros and cons of the cognitive computing utilization in distributed medical institutions. Moreover, the review highlights future research directions in the area such as data privacy concerns addressed, regulatory compliance, algorithm bias, and interdisciplinary collaboration between healthcare professionals and technologists. Through potential of cognitive computing in enhancing the precision and accuracy of diagnosis, improving patient healthcare, and reducing healthcare costs considerably in remote settings, this paper indeed provides useful information for researchers, practitioners, and stakeholders in the advancement of remote healthcare.

**Keywords** Virtual cognitive intelligence · Digital telehealth · Remote healthcare · Machine learning · Predictive analytics · Personalized treatment

---

R. Agarwal · A. K. Sinha · U. Dutta

Kalinga Institute of Industrial Technology, Deemed to Be University, Bhubaneswar, India  
e-mail: [21052208@kiit.ac.in](mailto:21052208@kiit.ac.in)

L. Wang (✉)

Xi'an Jiaotong-Liverpool University, Wuzhong District, Suzhou, China  
e-mail: [Stella.wang896@gmail.com](mailto:Stella.wang896@gmail.com)

B. Rathore

University of South Wales, Pontypridd, UK  
e-mail: [Bharati.rathore@southwales.ac.uk](mailto:Bharati.rathore@southwales.ac.uk)

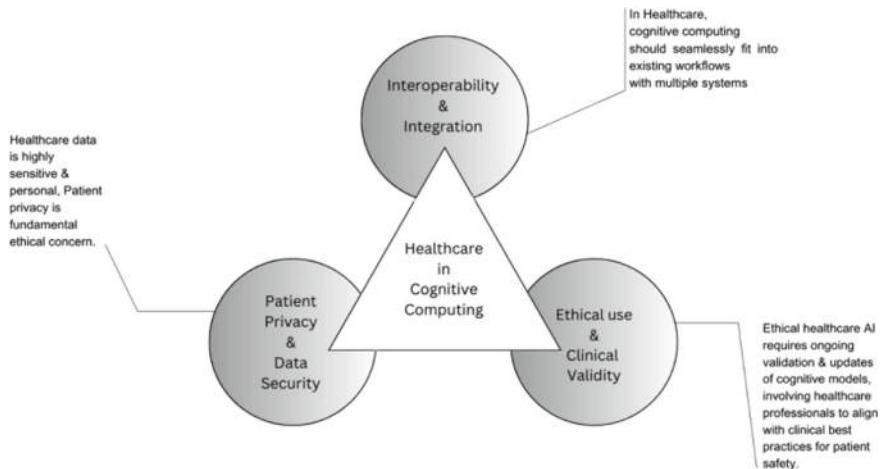
## 1 Introduction

Telehealth, which refers to using communication technologies, for healthcare delivery has become a crucial resource due to advances in communications, computer science, informatics, and medical technologies. In this field, a cognitive intelligence framework improves healthcare management by analyzing patient data. By employing algorithms and machine learning methods it helps create customized treatment plans and predictive analysis enhancing precision and simplifying processes. This approach allows for interventions and ongoing learning enhancing care and operational efficiency in the digital telehealth domain.

In Fig. 1, three essential features of cognitive computing in remote healthcare is shown and telecommunication advancements have made remote healthcare a crucial tool in addressing disparities in healthcare access, particularly in underprivileged and isolated regions. Nevertheless, providing quality services in isolated areas remains challenging as it is characterized poor infrastructure, fragmented healthcare systems and insufficient supply of specialized medical personnel. Usage of the cognitive computing technologies has the potential of improving the delivery methods of the telemedicine based form of healthcare. The human cognitive computing is only a subset of the artificial intelligence (AI) technology that was designed to mimic the human cognitive functions by using various technologies which are the natural language processing (NLP), data analytics (DA), and the machine learning (ML). It is supported by these technologies that the systems are able to perceive, reason and divulge from the vast amounts of medical data achieve better decisions and improve the quality of the care. The sphere of remote healthcare sees the cognitive computing as an opportunity for the improvement of the delivery of the care by means of the diagnosis support and remote monitoring as well as the use of the treatment planning and the medical imaging analysis. This paper offers an outline on the application of cognitive computing in remote healthcare, summarizing past research, and pinpointing major challenges and opportunities. Through an analysis of the current state-of-the-art and a projection to future directions, this review seeks to explore the transformative potential of cognitive computing in reducing health disparities and improving access to quality care in remote and resource-limited settings. Crucially, this paper highlights the significance of consideration of ethical, regulatory, and technological aspects exploit the potential of cognitive computing in remote healthcare that is free from risks or disparities.

Thereafter, all programs should exclusively access the database through a volatile interface defined in an executable of a particular name.

1. Clinical Decision Support Systems (CDSS): Cognitive computing systems can analyze patient data remotely, thus helping physicians to precisely diagnose conditions, even in regions that lack specialized medical expertise.
2. Remote Monitoring: Cognitive Computing-based AI systems for remote monitoring can track patients' vital signs and health metrics, giving the health professionals real-time insights and timely interventions.

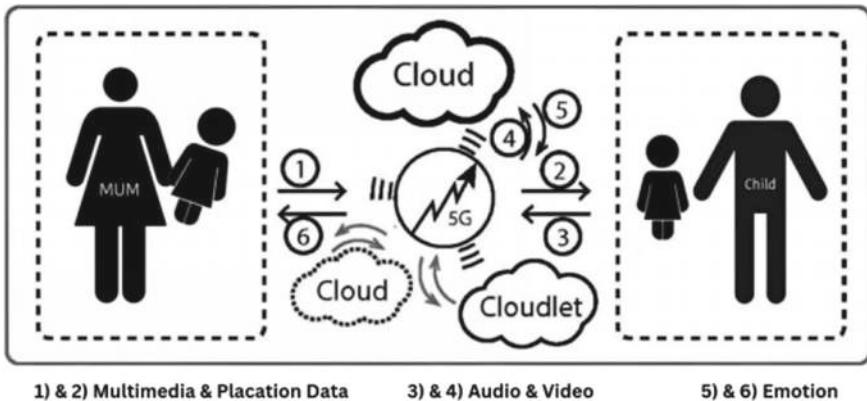


**Fig. 1** Pillars of remote healthcare

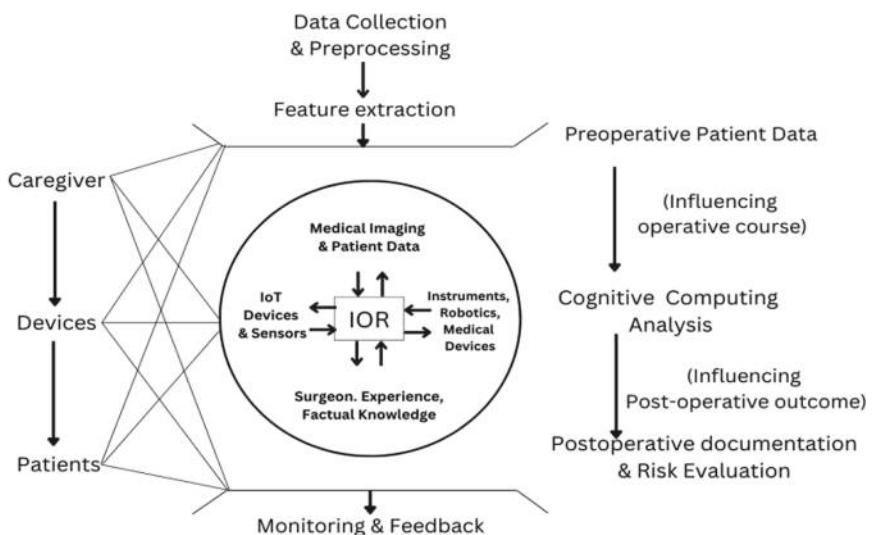
3. Personalized Treatment Plans: Using the cognitive computing algorithms the patient data analysis enables to provide remote personalized treatment plans based on personal health-related parameters, medical history, and medical equipment.
4. Virtual Health Consultations: Cognitive computing empowers teleconsultations by supplying support during remote engagement of patients by healthcare providers, thus boosting the quality and the availability of medical assistance.
5. Medical Imaging Analysis: In remote surgery, cognitive computing can participate at the stage of medical image analysis, e.g., MRI and CT, which can help surgeons make the right decisions and do the procedure remotely.
6. Remote Surgery: The constraints on time, geographical barriers, lack of proper medical care, and unavailability of esteemed doctors can be dangerous to a patient's life in critical cases. Remote surgery can be applied to overcome this problem (Fig. 2).
7. Remote Emotional Pacification: The system can be used to detect the emotional state of the individual and appropriate actions can be taken to neutralize the situation or better the mood of the subject. The model can be employed as an alternative for caretakers. Figure 3 depicts the mentioned situation where a mother can keep an eye on her baby using the system.

## 2 Literature Survey

In [1] authors state that the cognitive systems perform the functions of a brain and enhances decision making abilities to tap all the available information in numerous application areas. It delivers a model that responds by feeling, comprehending natural



**Fig. 2** Remote emotional pacification



[IOR-> Integrated OR]

**Fig. 3** Process framework of cognitive analysis in remote healthcare

language, and providing a response to the stimulus in a way similar to a human being not a traditional system that is programmable. Developing a custom chat box or search assistant to gain dialog with humans by interpreting queries and explaining data insights. The chapter also shows some of the problems of cognitive computing which are sometimes similar to what humans think. In [2] authors portray the progress of the sensors empowered systems and artificial intelligence algorithms, various

human healthcare applications stepped on. New system-oriented cognitive system paradigms are presented enabling various types of help, e.g., smart social interaction services insurance with sentimental connections and self-regulation motives. The paper gives the process of development and deployment of the cognitive-based healthcare applications in the cloud environment. Based that cloud service was being selected, proposed a cloud-based research on healthcare application architecture. Architecture to store, analyze and make predictions for biomedical big data. In [3] authors states that the mental disorders are characterized by impairments of cognitive control that affect mood, thinking, work performance, and functions such as physical abilities emotions and interpersonal interactions. This paper has a systematic review of computer approaches and technology that are appropriate for automatic recognition of mental diseases, the PRISMA model. Self-diagnostic methods, like questionnaires and rating scales are non-coherent and static missing the scope of variability inherent to mental disorders in their diversity. In [4] authors have found one of the key areas of research where a significant progress has been observed is the most modern technology is healthcare. The people who need healthcare services must go through the tedious process. Thus, in general the main issue is technology will focus on healthcare problems. The main objective of the project is to set up healthcare systems, which include three components. In [5] authors say that cognitive computing is transforming healthcares to such level where IT solutions that are automated can take care of community health issues and disease prevention. In [6] authors have shown in their study that cognitive computing is an intelligent system that mimics human behavior, such as on a large scale learning and deliberate reasoning. Organizing of corporate values have an effect on finance and investment companies, healthcare industries, traveling industry tourism interest in human life science education agriculture communications and network technology. Physician decision-making [7] is entirely transformed by cognitive computing where physicians are in better position to manage diseases and improve patient outcomes. As a result, physician decision-making is entirely transformed by cognitive computing where physicians are in better position to manage diseases and improve patient outcomes. This study uses cognitive computing to process a large amount of data, offer smart suggestions and enhance human-based decision making. In [8] authors have shown in their study that cognitive computing is revolutionizing studies and investigational research through improvement in efficiency and performance. In [9] authors have presented the paper dealing with one of the components that is the backbone of the whole operation, i.e., network. The paper proposes use of 5G network for faster transmission in the healthcare world where a split second decision could make the difference between life and death. In [10] authors discusses the ideal cognitive healthcare model integrated with IoT and cloud technologies to offer timely services through smart cities. The framework considers several modes of healthcare data. A higher accuracy is shown in experimental results of the deep learning method-based EEG pathology classification. It provides a neural network algorithm which gives advanced pathology detection results. Literature reviews, the proposed framework, experimental results and a conclusion section wrap up this paper. In [11] authors

brings light on the broad use of smart devices and sensors in many new application areas. IoT and cloud computing becoming mainstream over the past years demands our considerable attention. The system can be improved by tying a merger between telecom companies and healthcare providers to reserve lines for swift and secure communication. In [12] authors presents the lessons learned from corporally and speaks about global COVID-1 pandemic influence. It stresses on the need to invest in a vibrant non pharmaceutical intervention measure that can effectively stop transmissions of infectious viruses.

### 3 Proposed Model

In Fig. 3, the model shows how the cognitive computing analysis can redefine remote healthcare with its Smart Decision-Making feature. The whole model is a full-fledged structure of implementation in reality, that ability to use for real purpose by which people believe it as something feasible and reasonable nowadays at least before taking further decisions on this matter will be solved [13, 14]. Initially, we trigger the decision-making procedure by feeding it some input data such as collecting data from various sources: As it consistently gets worse, no one has any faith in the situation [15].

#### (1) Data Collection

Electronic Health Records (EHRs), patient histories, the sensor data such as wearables, IoT devices, medical literature, Government health databases as well as other relevant sources.

The second step guarantees the protection of information and compliance with privacy requirements using advanced encryption technologies, access controls as well as secure transfer methods. Comply with regulatory norms such as HIPAA and GDPR, besides performing audits of risks assessment to minimize potential breaches while safeguarding the confidential healthcare data [16–18].

#### (2) Data Preprocessing

Normalize data, eliminate outliers and missing values, standardizing formats; normalize variables. This guarantees data quality while making sure the accuracy of subsequent analysis.

Transform the data that is collected from healthcare into formats suitable for analysis by cognitive computing, ensuring proper formatting so as to apply machine learning and other forms of algorithms.

#### (3) Feature Extraction

Locate critical characteristics within preprocessed healthcare data that are crucial in informing decision-making procedures, with the aim of mining relevant information necessary for producing useful insights and recommendations.

#### (4) Cognitive Computing Analysis

Exploit state-of-the-art machine learning and cognitive algorithms to analyze the features that are detected in healthcare data, identifying meaningful patterns, and crucial insights for thoughtful decision-making. Use NLP to analyze unstructured healthcare data including clinical notes and research materials, distilling key findings that can inform the decision-making process in healthcare. Use pattern recognition and anomaly detection algorithms to recognize abnormal trends or patterns in healthcare data, thus identifying the deviations that are crucial for correct decision-making process within the medical environment. Through the healthcare data analysis with cognitive computing, meaningful insights and recommendations from cognitive computers can be obtained for health practitioners which help them to take more accurate decisions by providing necessary guidelines and directions. Offer detailed rationale for the recommendations derived from cognitive computing analysis; it encourages accountability and allows stakeholders to understand why certain suggestions were put forward. Present the gathered insights to healthcare professionals considering their competence and preferences in order for decision-making that reflects one's knowledge base with other peculiarities required from a healthcare professional. Establish a bi-directional communication channel between the cognitive system and healthcare providers allowing for information exchange, feedbacks in order to enhance decision-making processes within a healthcare environment.

Ensure patient involvement where applicable, allowing active participation of patients and considering their preferences, values, needs to achieve the concept of PCC in healthcare decisions. Making sure that information provided to patients is straightforward and accessible, so they can be understood and choose as relevant based on the origins where it given; languages from which for them are literate-numeracy capacity ones while preferred communication channel use. Rely on feedback loops to validate the precision and reliability of recommendations by taking into consideration reactions from stakeholders, as well as actual health system outcomes for continuous improvement. Validate outcomes against known healthcare benchmarks and standards upholding adherence with established protocols, using data in their compliance to guidelines that have been accepted as best practice for ensuring quality and safety of the decision-making process within medical settings. Combine cognitive analysis and insights with healthcare professionals' experience, patients preferences to synthesize different viewpoints that will help you make a decision which is in line with the health goals and priorities. In order to address the healthcare needs, take up what has been recommended in terms of treatment plans, interventions and further diagnostic tests and implement them so that at least a good decision-making process is followed which leads to desirable outcomes. There is an upsurge of scientific pursuits and industrial developments aiming at guiding and supervising surgeons not only by executing specific steps of surgical procedures but also by continuous observation and tutorship. Lately, there have emerged several commercial products termed "Integrated ORs." Nevertheless, these configurations use proprietary interfaces and control the equipment manually most of the time. Developing autonomous systems is not the main focus. Considering the existence of this gap, moving forward, "Cognitive Surgery" is introduced as a fully independent

surgical aide to surgeons by overlaying surgical training, patient data, and advanced data analytics into the Sensor OR framework.

- (5) Monitoring, Feedback, and End Users: Thus, within the course of a week in which it was given to him as he drew nearer toward his deathbed. Monitor the performance of patients on a regular basis and ask for their feedback as well as evaluate healthcare providers' input to check efficacy, modify interventions if required so that optimal results are achieved in patient management. Personalize data representation for different users including doctors, nurses, and patients with a view to ensure user-friendly interfaces that enhance effective interaction between the end use of cognitive system maximizing usability leading ultimately better healthcare decision making processes. The decision-making procedure has finished, but is still open to ongoing efficacy and education. This diagram gives a general description of the decision-making process, by using cognitive computing together with human skills and patient engagement in healthcare; it is an iterative approach to constant improvement. The specifics of implementation would also have to be adapted depending on the healthcare setting and needs.

## 4 Results and Discussion

The proposed model can be implemented using a virtual cognitive intelligence framework with the help of generative algorithms like K-Nearest Neighbors (KNN), neural networks, NLP, and deep learning to name a few. Programming of the algorithms can be achieved through R (statistical analysis and graphical overview tool) and Python incorporating modules like TensorFlow and Matplotlib.

In Fig. 4, the recorded data on the various number of patients collected from hospitals was fed to the program. The data containing noise, outliers and anomalies are removed followed by normalization. The data that makes it to the next step is the Valid Data. The data presented within our model is measured in gigabytes (GB). The graph shows size comparison of existing data and data after processing. The processed data is then made available for access and processing for further analysis.

Data filtering and error correction are critical processes applied before working with any dataset. The mentioned processes ensure the quality of data. Hence, preprocessing the data reduces the size in comparison with the original data, and it may be used for experimental result validation. According to the findings in Fig. 5, processed data is lower in size against our input data to the program. The decrease in data size helps to improve precision in analytics. The notable thing of the whole process is that both, the raw and processed data continue to grow with time highlighting the demand for good data handling techniques.

Though various data models can be used for processing the data, the accuracy of some models are not at par. It would be in our best interest to choose the one with the highest accuracy to not face any disparity and wrong analysis while working on the data. In Fig. 6, the graph gives us the accuracy of multiple models. Boosting has the highest accuracy among all and therefore for our case study, we choose to apply

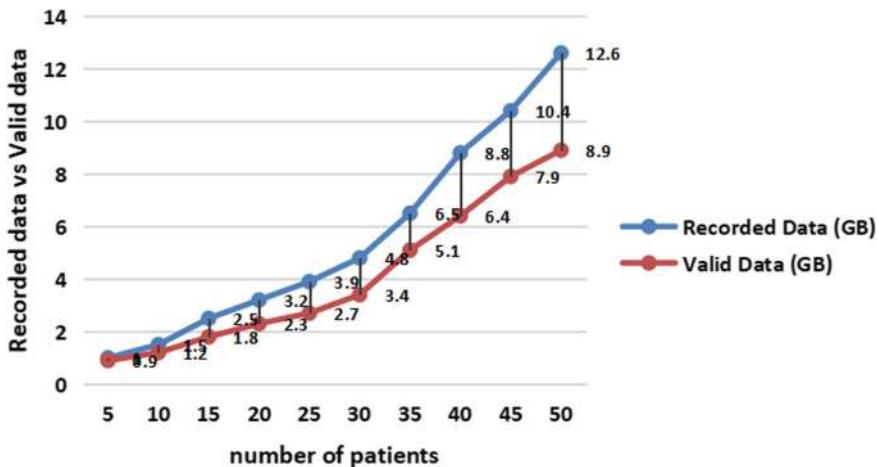


Fig. 4 Size comparison of existing data and after processing

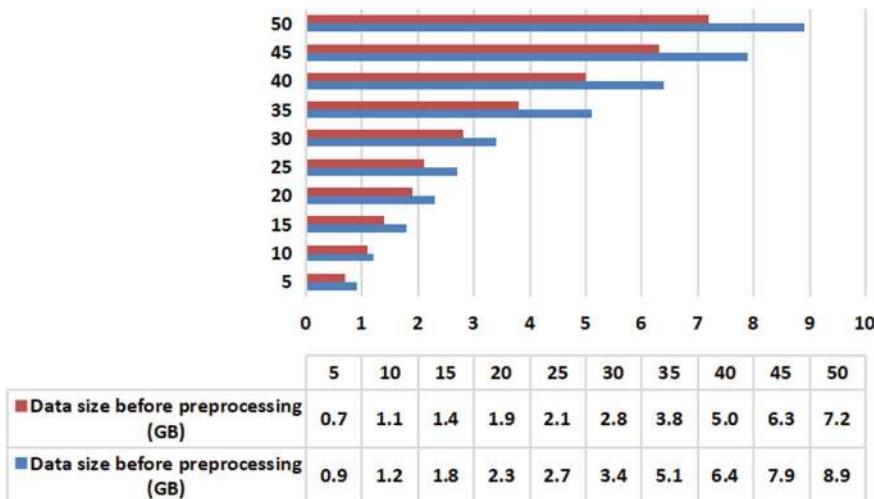
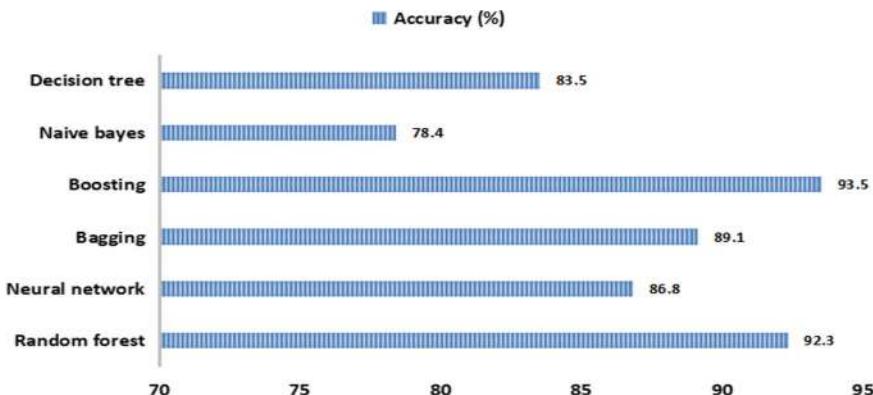


Fig. 5 Average compression in data before and after preprocessing

it. Boosting works on the idea of “wisdom of crowds” which is inherently better than any individual or standalone system. The model divides itself into several base learners as individuals in a crowd for the representation of human counterparts. Each learner focuses on one characteristic of data and produces corresponding bias. The cluster of insights from all the learners enable us to make a final decision. Thus, all the points make it a good model for working with data.



**Fig. 6** Accuracy analysis of the model

## 5 Conclusion

The use of remote healthcare can be the new wave to wellness and treatment. Individual attention to patients, accurate estimation, organized processes can cut expenditures and advance public health as a whole. The utilization of cognitive computing, complex algorithms, data mining, and other related technologies, healthcare can produce exceptional patient treatment. Remote surgery or telesurgery is an emerging technology in the field that uses cognitive computing. It utilizes both robotic technology and low latency network to connect patients and surgeons who are territorially diverged. This methodology can overcome the limitations of inaccessible high grade surgical care, unavailability of proficient surgeons, immediate attention to subjects with critical condition constrained by time. Despite the various advantages the technology can offer, it has its own shortcomings. There are considerations toward patient safety, initial cost of setup and maintenance, followed by legal and moral concerns. Privacy is a major distress as it governs the flow of data and is susceptible to network attacks which can complicate the ability to carry out telesurgery safely without posing a health hazard. The advantages that remote healthcare offer cross out the negatives. Improvements in individual health maintaining feasible costs can substantially impact society. The demand for proper healthcare is ever growing and remote healthcare can fit in the gaps bringing quality care for all.

## References

1. Sabarmathi KR, Leelavathi R (2019) Application of cognitive computing in healthcare. In: Cognitive social mining applications in data analytics and forensics, pp 265–272). IGI Global
2. Harini AS, Chakravorty C (2020) Cognitive cloud computing in healthcare. J Emerg Tech Inn Res

3. Singh J, Hamid MA (2022) Cognitive computing in mental healthcare: a review of methods and technologies for detection of mental disorders. *Cogn Comput* 14(6):2169–2186
4. Kumavat GD, Kumar S (2019) Health monitoring using edge cognitive computing based smart health care. *Int J Innov Tech Explor Eng* 9(1):2278–3075
5. Rastogi D, Tiwari V, Kumar S, Gupta PC (2022) Era of computational cognitive techniques in healthcare systems. *Cogn Intell Big Data Healthc*, pp 1–40
6. Aghav-Palwe S, Gunjal A (2021) Introduction to cognitive computing and its various applications. *Cogn Comput Hum Robot Interact*, pp 1–18
7. Behera RK, Bala PK, Dhir A (2019) The emerging role of cognitive computing in healthcare: a systematic literature review. *Int J Med Informatics* 129:154–166
8. Ahmed MN, Toor AS, O’Neil K, Friedland D (2017) Cognitive computing and the future of health care cognitive computing and the future of healthcare: the cognitive power of IBM Watson has the potential to transform global personalized medicine. *IEEE Pulse* 8(3):4–9
9. Chen M, Yang J, Hao Y, Mao S, Hwang K (2017) A 5G cognitive system for healthcare. *Big Data Cogn Comput* 1(1):2
10. Amin SU, Hossain MS, Muhammad G, Alhussein M, Rahman MA (2019) Cognitive smart healthcare for pathology detection and monitoring. *IEEE Access* 7:10745–10753
11. Chen M, Li W, Hao Y, Qian Y, Humar I (2018) Edge cognitive computing based smart healthcare system. *Futur Gener Comput Syst* 86:403–411
12. Jabbar MA, Shandilya SK, Kumar A, Shandilya S (2022) Applications of cognitive internet of medical things in modern healthcare. *Comput Electr Eng* 102:108276
13. Sahoo S, Mishra S, Brahma B, Barsocchi P, Bhoi AK (2024) SSO-CCNN: a correlation-based optimized deep CNN for brain tumor classification using sampled PGGAN. *Int J Comput Intell Syst* 17(1):1–18
14. Mishra S, Chaudhury P, Tripathy HK, Sahoo KS, Jhanjhi NZ, Hassan Elnour AA, Abdelmaboud A (2024) Enhancing health care through medical cognitive virtual agents. *Digital Health* 10:20552076241256732
15. Mishra S, Jena L, Mishra N, Chang HT (2024) PD-DETECTOR: a sustainable and computationally intelligent mobile application model for Parkinson’s disease severity assessment. *Heliyon* 10(14)
16. Pranjali P, Mallick S, Paul A, Mishra S, Bhardwaj I, Albuquerque VHCD (2024) Soil crops and nutrients forecasting using random forest model. In: AIP conference proceedings, vol 2919(1). AIP Publishing
17. Mishra S, Chakraborty S, Sahoo KS, Bilal M (2023) Cogni-Sec: a secure cognitive enabled distributed reinforcement learning model for medical cyber–physical system. *Internet Things* 24:100978
18. Mishra S, Volety DR, Bohra N, Alfarhood S, Safran M (2023) A smart and sustainable framework for millet crop monitoring equipped with disease detection using enhanced predictive intelligence. *Alex Eng J* 83:298–306

# Quality Detection Model of Nutmeg (*Myristica Fragrans Houtt*) Using You Only Look Once (YOLO)



Manuel Soares Dos Reis Pacheco, Hadiyanto Hadiyanto,  
and Ridwan Sanjaya

**Abstract** This research proposes a new approach to detect nutmeg (*Myristica fragrans Houtt*) quality using the "You Only Look Once" (YOLO) object detection technique. This method automatically identifies nutmeg quality based on visual attributes like color, size, and wholeness. Nutmeg image data are collected and processed using the YOLO model to distinguish between high and low-quality nutmegs. Experimental results show that this approach successfully provides a high accuracy rate of 95.6% in detecting nutmeg quality, paving the way for developing an efficient and fast nutmeg quality classification automation system.

**Keywords** Object detection · Computer vision · Image processing · Nutmeg · Quality · YOLO

## 1 Introduction

One of the most important spice ingredients in the world's food and beverage sector is nutmeg (*Myristica fragrans Houtt*). Nutmeg also has significant economic value apart from its use as a flavor additive. Indonesia is one of the most famous countries in the world for producing and exporting nutmeg. The spice plant nutmeg (*Myristica fragran haitt*) has a significant economic value, particularly the parts and mace that can be used to make nutmeg oil, also known as oleum myristicae, oleum myrist, or myristica oil [1, 2]. Indonesian nutmeg has a high selling value on the world market and is multipurpose since every part of the plant can be used in various

---

M. S. Dos Reis Pacheco (✉) · Hadiyanto Hadiyanto  
Diponegoro University, Semarang, Indonesia  
e-mail: [nacel1983@gmail.com](mailto:nacel1983@gmail.com)

Hadiyanto Hadiyanto  
e-mail: [hadiyanto.chm@undip.ac.id](mailto:hadiyanto.chm@undip.ac.id)

R. Sanjaya  
Soegijapranata Catholic University, Semarang, Indonesia  
e-mail: [ridwan@unika.ac.id](mailto:ridwan@unika.ac.id)

industry types [3]. The increasing supply of nutmeg products to the world market has forced export destination countries to set higher and more stringent quality standards, including the European Union. Only nutmeg products that meet sanitation standards, i.e., those accompanied by a certificate of nutmeg sample analysis results containing information on mycotoxin content consisting of aflatoxin and ochratoxin contents, are permitted to enter the area [4]. Even though nutmeg is a leading commodity in the plantation sector, its quality assessment is often subjective. It depends on the individual grower's experience, which can cause differences in quality assessment between one farmer and another. Since the assessment or sorting mechanism is still based on the individual farmer's experience, farmers or collectors need help to optimally meet the massive demand for nutmeg exports, resulting in a decline in selling prices [5]. The quality of a product can be seen from several variables. The variables that influence quality are color and shape [6]. These two variables are considered one of the most important parameters when selecting a product. The increase in global nutmeg production and the strict regulations imposed by nutmeg-importing countries are inseparable, which creates more choices for obtaining high-quality nutmeg products. Several researchers have studied to identify certain fruits, using image processing based on texture, color, and shape [7], applying the YOLO model to identify various types of vegetables and fruits [8], and automatic tomato detection using the Yolov5 model with CNN [9], using artificial intelligence for classification of nutmeg ripeness [10], and implementing of an image segmentation algorithm for identifying ripe nutmeg [11]. Furthermore, this research aims to develop a model for determining the maturity level of nutmeg seed quality based on the Indonesian National Quality Standards, namely Calibrated Nutmeg (CN), ABCD Average, Rimpel (Shrivel), and Broken Wormy Punky (BWP) quality classes using You Only Look Once (YOLO).

## 2 Theoretical Background

### 2.1 Computer Vision

Computer vision is a computer science that focuses on understanding and analyzing visual content from the real world [12]. The main goal is to develop algorithms and technologies that enable computers to process, analyze, and understand images and videos in a human-like way [13]. Computer vision is related to several fields, namely image processing and machine vision. The various techniques and applications that span these three areas share significant similarities.

## 2.2 Object Detection

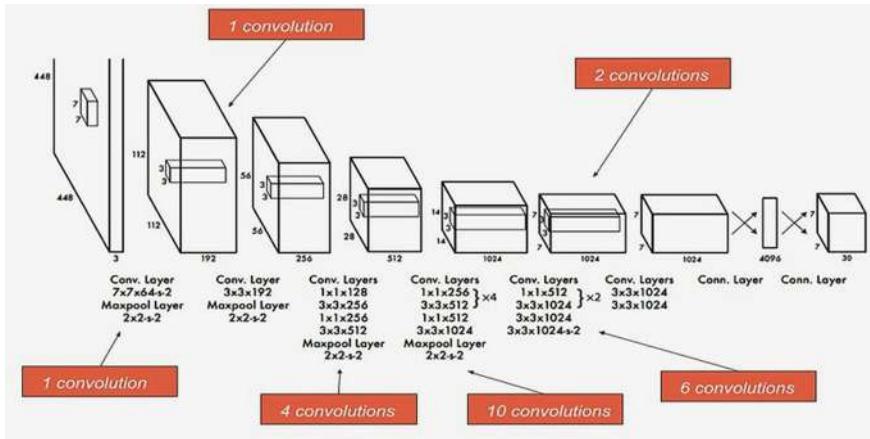
Object detection has recently become a major focus as one of the most crucial and challenging areas in computer vision. Over the past twenty years, advancements in object detection technology have rapidly progressed, significantly impacting the field of computer vision [14]. Advances in hardware and current processing resources are occurring rapidly and innovatively [15]. Object detection is one of the tasks in image processing and computer vision, which aims to identify the location and type of objects in images or videos [16]. Object detection systems build models for object classes from a training data set. Object detection methods can be used for various applications, including security surveillance, facial recognition, medical image analysis, robotics, and more.

## 2.3 Feature Extraction

Feature extraction in image processing is identifying and extracting relevant information or important features from image data. The main goal of feature extraction is to change the representation of an image into a simpler and more compact form while still retaining information that is important for the task to be solved, such as classification, object detection, or pattern recognition. The main challenge of feature extraction is learning and extracting knowledge from that data to make correct decisions [17].

## 2.4 You Only Look Once (YOLO)

YOLO, which stands for You Only Look Once, is an advanced real-time object detection system [18]. This system was first introduced by Joseph Redmon et al. [19]. The architecture of the YOLO model is shown in Fig. 1. This model is famous for its accuracy and speed. YOLO uses artificial neural networks trained with extensive datasets of images and videos. This dataset contains images and videos that have been labeled, where each label indicates the objects in the image or video. YOLO bounding box labels start with the class ID number, followed by the box coordinates normalized between 0 and 1. It is important to note that the  $x$  and  $y$  coordinates here are the centers of the bounding box. They have chosen this to make bounding boxes scalable regardless of image size. The YOLO target identification algorithm's compact model size and quick computation time are its key features. Because of its simple structure, YOLO may use a neural network to display the position and category of bounding boxes directly. YOLO is able to pick up highly transferable traits to different domains [20]. Yolo has 24 convolutional layers, alternating  $1 \times 1$  convolutional layers, reducing the feature space of the previous layer, and then continuing with two fully connected layers to produce the final tensor. Since its



**Fig. 1** YOLO Full Model Architecture [22]

launch in 2015, the YOLO object detection variant has overgrown, with the latest release of YOLO-v9 in February 2024 [21].

## 2.5 Quality of Nutmeg Seeds According to Indonesian National Standard (SNI) 01-0006-1993

These are round or oval, 20–40 mm long seeds from the fruit of the *Myristica* spp. plant that have been dried and peeled off the shell, per SNI 01-0006-1993 [23]. As shown in Fig. 2, nutmeg is made up of nutmeg mace, shell, and fruit flesh. The four (four) quality categories for Indonesian nutmeg are Calibrated Nutmeg (CN), ABCD Average, Rimpel (Shrivel), and BWP. Additionally, Table 1 shows the SNI 01-0006-1993 nutmeg quality classification.



**Fig. 2** Component of nutmeg fruit [24]

**Table 1** Quality of nutmeg classification based on SNI 01-0006-1993

Characteristics	Calibrated nutmeg (C N)	ABCD average	Shrivel	BWP
Weight (gram)	4.11–8.33	Relative density	Relative density	Lighter weight
Shape	Smooth surface without wrinkled	All around and less wrinkled	Not all around and wrinkled	Wrinkled
Broken	No	No	No	Yes
Contaminated (%)	< 10	< 10	< 10	N/A

## 2.6 Support Vector Machine

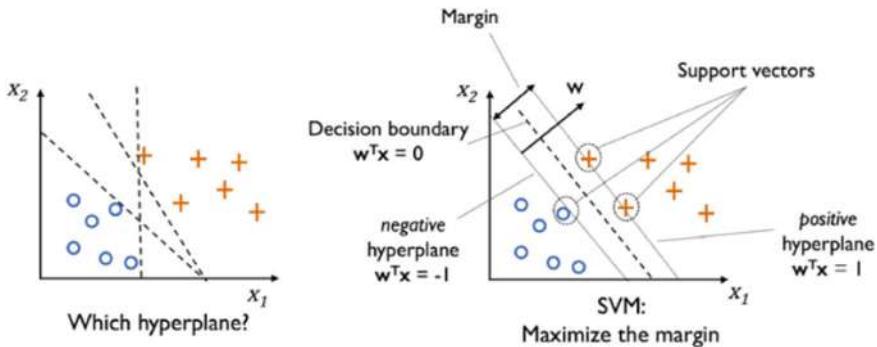
Classification using the support vector machine (SVM) is one of the popular approaches in machine learning to separate two or more classes based on the features in the dataset [25, 26]. SVM is a supervised learning technique commonly employed for classification tasks [27]. SVM aims to find the hyperplane with the most significant margin between different classes in the data [28]. A hyperplane is a decision boundary that divides the feature space into two regions for other classes. SVM takes data as input in the form of numerical vectors. Each vector represents one object in feature space, where each dimension in the vector may represent different features of that object. The linear hyperplane equation is written in Eq. 1 below:

$$w^T x + b = 0 \quad (1)$$

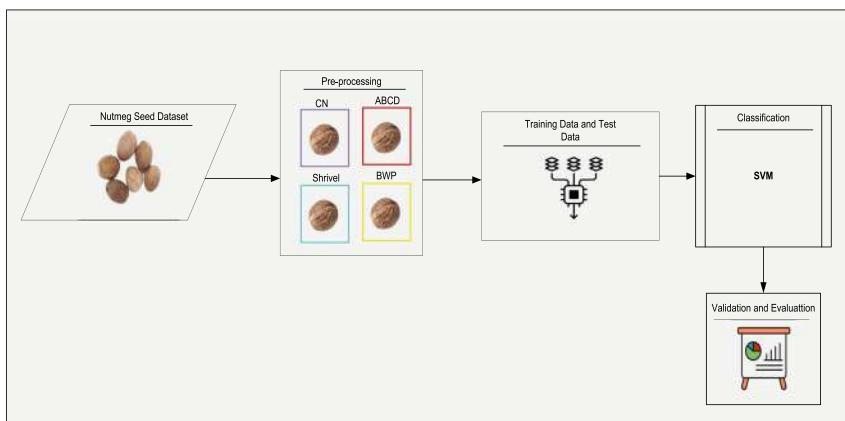
The hyperplane identified by SVM is shown in Fig. 3; it is located in the middle between two classes. This shows that the distance between the hyperplane and the object data from the adjacent (outermost) class, which is marked with an empty and positive circle, varies. The outermost object data closest to the hyperplane in SVM is called the support vector. The support vector is the most difficult data to classify because its position almost intersects with other classes. SVM performance can be further improved with kernel functions [29, 30]. This enables the development of nonlinear, higher-dimensional models. In case the problem is nonlinear, the kernel transforms it into a linear problem in the resulting higher-dimensional space by introducing new dimensions to the original data.

## 3 Methodology

This research started with data collection, pre-processing, model training using training and test data, model validation, and performance evaluation. Figure 4 shows the stages of the proposed research and the stages.



**Fig. 3** Hyperplane that separates the two classes positive (+1) and negative(-1)



**Fig. 4** Research stages

### 3.1 Dataset Collection

Good quality and representation of the dataset are crucial to the success of the nutmeg quality detection model using YOLO. A thorough and careful data collection process will ensure that the resulting model can provide accurate and reliable results in detecting the quality of nutmeg seeds.

The data collection stages in this research were carried out as follows:

1. The primary data related to the image of nutmeg seeds of various qualities would be the main focus. This data could be obtained from several sources, such as direct photos from multiple sources of nutmeg seeds in the field.
2. The existing secondary data, for example, were from a general existing database of nutmeg images.

### **3.2 Pre-processing**

The data pre-processing stage in this research is essential to prepare the raw data before it is used for model training and testing. The following are the typical stages in data pre-processing:

1. Data cleaning

Eliminating irrelevant data: Irrelevant data that does not match the research objectives is removed, e.g., blurry images or unrelated to nutmeg.

2. Data augmentation

Data augmentation aims to increase a dataset's diversity by adding variation to existing data.

3. Labeling

Each image of cleaned and processed nutmeg seeds must have a label or annotation corresponding to the seeds' quality. These labels must be adapted to the format required by the YOLO detection algorithm.

### **3.3 Division of Training and Test Data**

Dividing data into training and test data is critical in developing a machine learning model. The function of the training data is to form a classification model, while the test data has a function as a test in evaluating the quality of the results. The composition of training data and test data that would be used in this research was 80:20.

### **3.4 Classification**

The next stage was the classification process after dividing the training and test data. The classification process in this research used the support vector machine (SVM) method. SVM was used to classify the quality of nutmeg seeds based on features extracted from images using YOLO as an object detector. SVM was chosen since it could handle complex data and found optimal linear or nonlinear separators between different classes [31]

### 3.5 Validation and Evaluation

The evaluation aimed to provide a more in-depth picture of how well the model could predict or classify data. This research used a confusion matrix to measure accuracy, precision, and recall.

## 4 Results and Discussion

Tests were carried out to evaluate how well the model could recognize and differentiate between types of nutmeg seed quality, namely Calibrated Nutmeg (CN), ABCD Average, Rimpel (Shrivel), and BWP. This test involved 400 samples of nutmeg image data to test the model with each class. Table 2 presents the findings from testing and assessing the quality of nutmeg seeds.

Table 2 shows the results of our test image capture in detecting the quality of nutmeg seeds using YOLO. The results above show that the model created produced

**Table 2** Results of testing and evaluation of the quality of nutmeg seeds

Image sample	Actual label	Detection results	Accuracy	Precision	Recall
	Calibrated Nutmeg		<b>0.956</b>	<b>0.835</b>	<b>0.925</b>
	ABCD		0.911	0.776	0.810
	Shrivel		0.947	0.821	0.917
	BWP		0.941	0.819	0.886

the highest accuracy of 0.956 or 95.6% for the Calibrated Nutmeg (CN) quality nutmeg type. However, in the detection process carried out, there were still types of nutmeg images that did not match the prediction results, namely the ABCD Average nutmeg image was detected by the BWP nutmeg image, whereas the ABCD Average nutmeg image detected the BWP nutmeg image. This could happen since the features the model considers essential to differentiate between CN, BWP, and ABCD Average must be clarified or similar between categories. In addition, SVM relies on data separation based on extracted features [32]. If the extracted features need to be more discriminative or reflect significant differences between different nutmeg classes, the SVM may need help classifying them correctly.

## 5 Conclusion

This paper presents a nutmeg quality detection model that is capable of classifying seeds into four types of nutmeg quality in Indonesia with an accuracy level of 0.956% or 95.6% for the Calibrated Nutmeg (CN) type of nutmeg quality image, 0.911 or 91.1% for nutmeg quality ABCD Average, 0.947 or 94.7% for the Shrivel type, and 0.941 or 94.1% for the BWP nutmeg image. The use of YOLO and SVM in this study shows a complementary approach to detecting the quality of nutmeg seeds. YOLO is used for fast and accurate object detection in images, while SVM is used for further classification based on the extracted features. Recommendations for further research are to explore hyperparameters such as learning rate, batch size, and grid size on YOLO to get optimal results. Iterative experiments and cross-validation can do this optimization.

## References

1. Juliani Purba H, Supriadi Yusufi E, Hestina J (2021) Performance and competitiveness of indonesian nutmeg in export market. E3S Web Conf 232:1–13. <https://doi.org/10.1051/e3sconf/202123202018>
2. Ashokkumar K, Simal-Gandara J, Murugan M, Dhanya MK, Pandian A (2022) Nutmeg (*Myristica fragrans* Houtt.) essential oil: a review on its composition, biological, and pharmacological activities. Phyther Res 36(7):2839–2851. <https://doi.org/10.1002/ptr.7491>
3. Dharmaputra OS, Ambarwati S, Retnowati I, Nurfadila N (2022) Postharvest Quality improvement of nutmeg (*Myristica fragrans*). Biotropia (Bogor) 29(3):185–192. <https://doi.org/10.11598/btb.2022.29.3.1393>
4. Wahidin D, Purnhagen K (2018) Improving the level of food safety and market access in developing countries. Heliyon 4(7):e00683. <https://doi.org/10.1016/j.heliyon.2018.e00683>
5. Nasution IS, Gusriyan K (2019) Nutmeg grading system using computer vision techniques. IOP Conf Ser Earth Environ Sci 365(1). <https://doi.org/10.1088/1755-1315/365/1/012003>
6. Chitturi R, Carlos Londono J, Alberto Amezquita C (2019) The influence of color and shape of package design on consumer preference: the case of orange juice. Int J Innov Econ Dev 5(2):42–56. <https://doi.org/10.18775/ijied.1849-7551-7020.2015.52.2003>

7. Pratibha S, Abhishek D, Snehlata M (2022) Fruit recognition using image processing, i-manager's. *J Image Process* 9(3):10. <https://doi.org/10.26634/jip.9.3.19047>
8. Latha RS et al (2022) Fruits and vegetables recognition using YOLO. In: 2022 International conference on computer communication and informatics (ICI), pp 1–6. <https://doi.org/10.1109/ICCCIS4379.2022.9740820>
9. Tsai FT, Nguyen VT, Duong TP, Phan QH, Lien CH (2023) Tomato fruit detection using modified Yolov5m model with convolutional neural networks. *Plants* 12(17). <https://doi.org/10.3390/plants12173067>
10. Qisthi IB, Siswono H (2024) Classification of nutmeg ripeness using artificial intelligence. *IAES Int J Artif Intell* 13(2):2441–2450. <https://doi.org/10.11591/ijai.v13.i2.pp2441-2450>
11. Jerusalin CJ, Lenin FA, Mersheba FL, Dani D (2024) Identification of mature Nutmeg using colour space segmentation algorithm. *Curr Agric Res J* 11(3):727–739. <https://doi.org/10.12944/carj.11.3.04>
12. Potter K (2024) Computer vision and image recognition techniques. *Artic J Sci Conf Proc*. [Online]. Available: <https://www.researchgate.net/publication/379035447>
13. Matsuzaka Y, Yashiro R (2023) AI-Based computer vision techniques and expert systems. *AI* 4(1):289–302. <https://doi.org/10.3390/ai4010013>
14. Zou Z, Chen K, Shi Z, Guo Y, Ye J (2023) Object detection in 20 years: a survey. *Proc IEEE* 111(3):257–276. <https://doi.org/10.1109/JPROC.2023.3238524>
15. Padilla R, Netto SL, Da Silva EAB (2020) A survey on performance metrics for object-detection algorithms. *Int Conf Syst Signals, Image Process* pp 237–242. <https://doi.org/10.1109/IWSSIP48289.2020.9145130>
16. Amit Y, Felzenszwalb P, Girshick R (2020) Object detection. *Comput Vis*, pp 1–9. [https://doi.org/10.1007/978-3-030-03243-2\\_660-1](https://doi.org/10.1007/978-3-030-03243-2_660-1)
17. Corke P (2022) Image feature extraction. *Springer Tracts Adv Robot* 142:157–202. [https://doi.org/10.1007/978-3-030-79175-9\\_5](https://doi.org/10.1007/978-3-030-79175-9_5)
18. Diwan T, Anirudh G, Tembhurne JV (2023) Object detection using YOLO: challenges, architectural successors, datasets and applications. *Multimed Tools Appl* 82(6):9243–9275. <https://doi.org/10.1007/s11042-022-13644-y>
19. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: Unified, real-time object detection. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pp 779–788. <https://doi.org/10.1109/CVPR.2016.91>
20. Jiang P, Ergu D, Liu F, Cai Y, Ma B (2021) A review of yolo algorithm developments. *Procedia Comput Sci* 199:1066–1073. <https://doi.org/10.1016/j.procs.2022.01.135>
21. Wang C-Y, Yeh I-H, Liao H-YM (2024) YOLOv9: learning what you want to learn using programmable gradient information, [Online]. Available: <http://arxiv.org/abs/2402.13616>
22. Shenoda M (2023) Real-time object detection: YOLOv1 Re-implementation in PyTorch, pp 1–4, [Online]. Available: <http://arxiv.org/abs/2305.17786>
23. Dinar L, Suyantohadi A, Fajar FMA (2013) Kajian standar Nasional Indonesia Biji Pala. *J Stand* 15(2):83. <https://doi.org/10.31153/j.s.v15i2.111>
24. Paulus E, Suryani M (2019) Image analysis for a smart machine of nutmeg sorting. *J Phys Conf Ser* 1196(1). <https://doi.org/10.1088/1742-6596/1196/1/012059>
25. Sun X, Liu L, Wang H, Song W, Lu J (2016) Image classification via support vector machine. In: Proceedings 2015 4th International Conference Computing Science Network Technology ICCSNT 2015, ICCSNT, pp 485–489. <https://doi.org/10.1109/ICCSNT.2015.7490795>
26. Chandra MA, Bedi SS (2021) Survey on SVM and their application in image classification. *Int J Inf Technol* 13(5). <https://doi.org/10.1007/s41870-017-0080-1>
27. Mhetre P, Bapat MS (2015) Classification of teaching evaluation performance using support vector 4(6):37–39
28. Srivastava DK, Bhambhu L (2010) Data classification using support vector machine. *J Theor Appl Inf Technol* 12(1):1–7
29. Niranjan Kumar P (2024) SVM-based classifier for early detection of Alzheimer's disease. *Educ Adm Theory Pract* 30(5):1120–1131. <https://doi.org/10.53555/kuey.v30i5.3022>

30. Elahifasaee F (2022) Optimized SVM using AdaBoost and PSO to classify brain images of MR. In: 2022 International conference on machine vision and image processing (MVIP), pp 1–5. <https://doi.org/10.1109/MVIP53647.2022.9738549>
31. Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, Lopez A (2020) A comprehensive survey on support vector machine classification: applications, challenges and trends. Neurocomputing 408:189–215. <https://doi.org/10.1016/j.neucom.2019.10.118>
32. Chiong R, Fan Z, Hu Z, Chiong F (2021) Using an improved relative error support vector machine for body fat prediction. Comput Methods Programs Biomed 198:105749. <https://doi.org/10.1016/j.cmpb.2020.105749>

# Development of New Monolithic Zeolite Chitosan Hydrogel with Preliminary Adsorption Studies to Remove Organic Dyes from Aqueous Solutions



Ghassan Abbas Alwan, Hamida Idan Salman, and Asim A. Balakit

**Abstract** One of the methods used to tackle pollution problems is the development of efficient materials that can act as adsorbents or filters to remove organic pollutants from aqueous solutions. In the present work, a new monolithic zeolite chitosan hydrogel (ZCS) has been synthesized by freeze–thaw technique. FT-IR spectroscopy, SEM, and XRD analysis have characterized the synthesized hydrogel, the data confirmed that ZCS is porous, composed of folded sheets with zeolite particles on the surface of the sheets. The water uptake percentage by ZCS (acid and base treated samples) was found to be in the range 84–95%. Preliminary adsorption studies showed that ZCS is efficient adsorbent to remove congo red dye, and with moderate efficiency for the elimination of the colour methylene blue in aqueous solutions.

**Keywords** Zeolite · Chitosan · Hydrogel · Adsorption · Organic dyes

## 1 Introduction

Approximately half of the hazardous wastewater generated by the paper, cosmetics, textile, plastic, food, and pharmaceutical industries originates from dye [1–3]. The detrimental and obstructive impact of dyes on aquatic environments is a significant issue [4, 5]. Dyeing is a complex and durable procedure due to the inert nature of the produced dye and its low concentration in water, which causes difficulties in removing

---

G. A. Alwan (✉) · H. I. Salman

Department of Chemistry, University of Kerbala, Kerbala, Iraq

e-mail: [ghassan.abbas@s.uokerbala.edu.iq](mailto:ghassan.abbas@s.uokerbala.edu.iq)

H. I. Salman

e-mail: [hamida.idan@uokerbala.edu.iq](mailto:hamida.idan@uokerbala.edu.iq)

A. A. Balakit

College of Pharmacy, University of Babylon, Babylon, Iraq

e-mail: [phar.asim.balakit@uobabylon.edu.iq](mailto:phar.asim.balakit@uobabylon.edu.iq)

it [6]. Adsorption has emerged as a prominent technique for extracting pigment from water-based solutions over the past decade [7]. Due to its high efficacy in eliminating organic dye molecules, adsorption was commonly employed for colour removal [8, 9]. Furthermore, it was simple to administer and could operate well even at low levels of dye concentrations. Chitosan hydrogels have numerous functional groups, such as hydroxyl and amine groups, which enable them to effectively eliminate pollutants from water [10]. Zeolite's high surface area and ion exchange capacity provide a three-dimensional matrix of linked pores, distinguishing it from other naturally occurring adsorbents. A zeolite is a microporous material composed of crystalline aluminosilicate. The efficiency of dye removal from wastewater with zeolite is 94% [11–13]. We specialize in the production of a zeolite chitosan hydrogel. A straightforward technique for synthesizing hydrogel is freeze–thaw synthesis. This hydrogel (ZCS) may be advantageous for optimizing organic dye removal in water filters.

## 2 Materials and Methods

Chitosan (M.wt. 3800–20,000 daltons), zeolite mordenite, glutaraldehyde (25%), congo red, methylene blue, and acetic acid were obtained from commercial origins and used as received without further purification.

### 2.1 *Synthesis of the Zeolite—Chitosan Hydrogel (ZCS)*

A solution was made by dissolving 1.0 grammes of chitosan in water-based acetic acid. The initial solution, containing 1% (50.0 ml), was constantly agitated at ambient temperature. Under ambient conditions, 0.75 grammes of zeolite were combined with 20 ml of water and vigorously agitated to produce solution 2. Option Three: To produce glutaraldehyde, dilute 1.0 mL of the 25% solution with 20.0 mL of purified water. Following the combination of solutions 1 and 2, solution 3 was introduced and vigorously mixed for a duration of 30 s. The liquid was thereafter put into 1 cm syringes and allowed to thoroughly dry for a duration of 24 h. Prior to being melted at room temperature, the hydrogel samples were stored at –40 °C for a duration of 48 h.

Before being stored until required, samples were treated with a 0.1 M NaOH or 0.01 M HCl solution, were washed with distilled water, and then stored.

## 2.2 Scanning Electron Microscopy (SEM)

Scanning electron microscopy (SEM) images were obtained with a Quanta 450 Field Emission Inverter (FEI) equipped with a voltage of 12.5 kV, a spot size of 1.5 mm, and a magnification range of 2.3–1355 X.

## 2.3 Fourier-Transform Infrared Spectroscopy

Fourier-transform infrared (FT-IR) spectra were obtained for the hydrogel and chitosan utilizing the ATR technique using a Bruker Tensor II instrument.

## 2.4 X-ray Diffraction (XRD)

The XRD analyses were performed on ADX-2700 instrument.

## 2.5 Water Uptake Capacity (WU) of ZCS

Samples of the hydrogel were left to dry at room temperature until constant weight was reached. The weight of the dry samples ( $W_{\text{dry}}$ ) were recorded, after immersing in water with different pH values for 24 h, the weight of the wet samples ( $W_{\text{wet}}$ ) was noted, and the water uptake capacity (WU) was determined using the formula below. Samples were cleaned with tissue paper to remove surface water.

$$\text{WU} = \left( \frac{W_{\text{wet}} - W_{\text{dry}}}{W_{\text{dry}}} \right) \times 100 \quad (1)$$

## 2.6 Adsorption Study

After stirring a 0.05 gramme hydrogel ZCS in 30.0 millilitres of dye solutions containing methylene blue (MB) and congo red (CR) (40 parts per million) for 90 min, the resulting solutions were analysed for absorbance at wavelengths of 664 nm and 498 nm using a Shimadzu 1800 UV-Visometer. The removal percentage values ( $R\%$ ) were calculated according to the following equation:

$$R\% = \frac{C_o - C_e}{C_o} \times 100\% \quad (2)$$

where  $C_o$  and  $C_e$  are the initial and equilibrium dye concentrations, respectively.

### 3 Results and Discussion

#### 3.1 Synthesis and Characterization of ZCS

To get porous monolithic polymer with the zeolite particles distributed on its sheets surfaces(ZCS), freeze and thaw technique was implemented [14], solution of chitosan was mixed with the suspension of zeolite, then the cross-linking was achieved by adding glutaraldehyde, the formed hydrogel was immediately casted and left for ageing, then after freezing at  $-40^{\circ}\text{C}$ , the monoliths were thawed. Acidic and basic treatments were done to study the effect of that on the adsorption of the organic dyes (CR and MB) on the surface of the ZCS. Figure 1 illustrates the synthesis steps.

#### 3.2 Fourier-Transform Infrared Spectroscopy

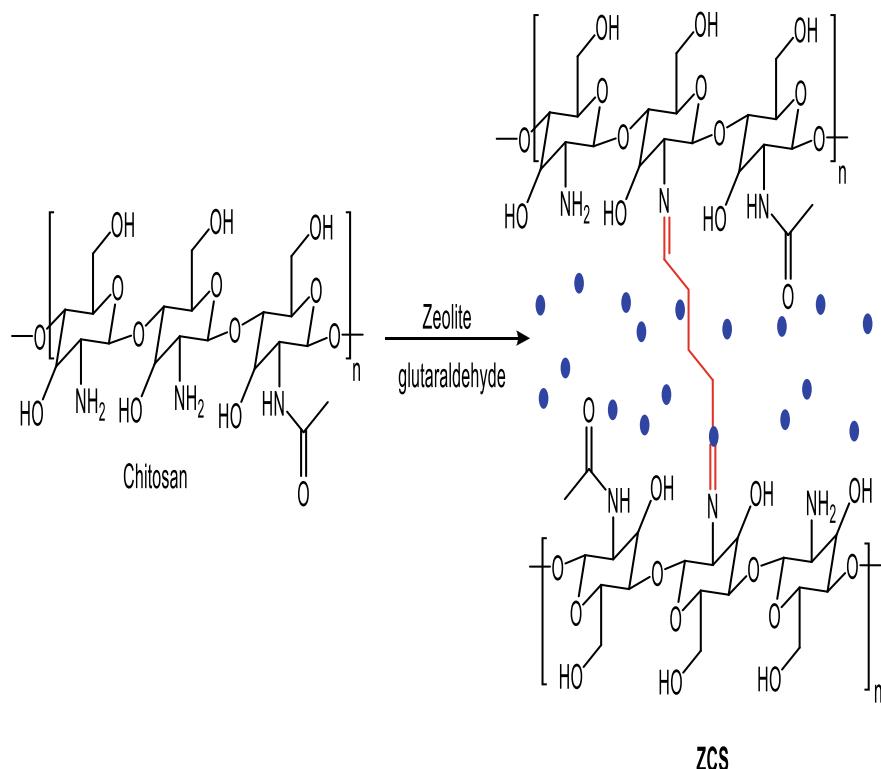
In Fig. 2, the FT-IR spectra of ZCS and chitosan are shown. FT-IR spectra of ZCS exhibit absorption bands at  $3316\text{ cm}^{-1}$  for the O–H bond in  $\text{H}_2\text{O}$ ,  $1647\text{ cm}^{-1}$  for the C = N bond (imine connection between chitosan and glutaraldehyde),  $1029\text{ cm}^{-1}$  for the Si–O bond in the zeolite (contributing to the three-dimensional silica phase), and  $625$  and  $554\text{ cm}^{-1}$  for the Si–O–Si bending vibration.

#### 3.3 SEM Analysis

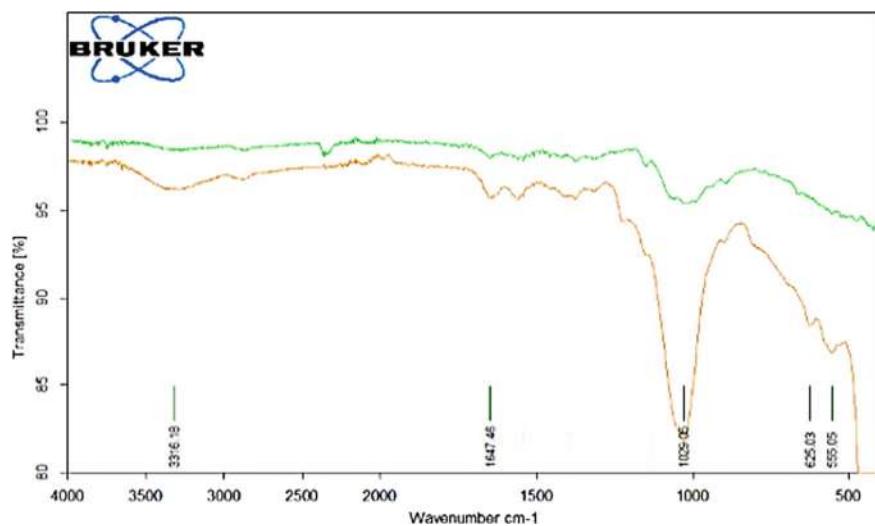
Analysis of the surface cross-section of the ZCS piece was conducted using a scanning electron microscope. The micrographs in Fig. 3 exhibit two degrees of magnification. The roughness of the surface is attributed to its composition of folded polymeric sheets that are embedded with zeolite particle.

#### 3.4 XRD Analyses

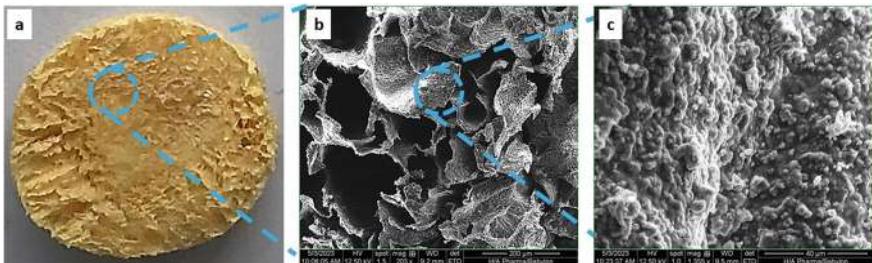
The XRD spectrum of ZCS is shown in Fig. 4. Table 1 includes the obtained data. Scherrer's equation was used to determine the ZCS polymer's average crystallite size.



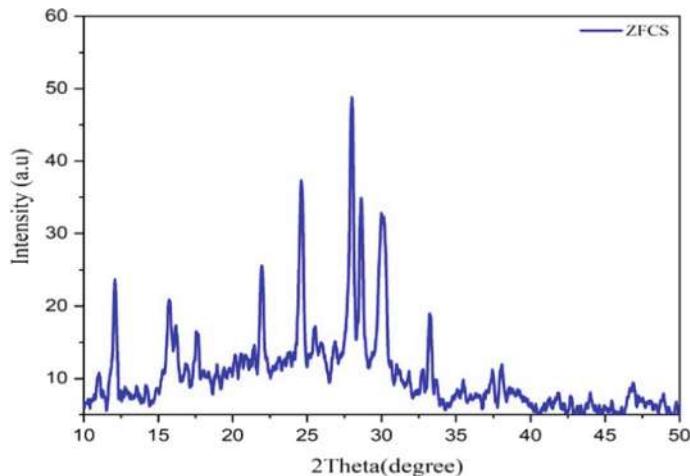
**Fig. 1** Synthesis of ZCS



**Fig. 2** FT-IR spectra of chitosan (green) and ZCS (red)



**Fig. 3** Cross-section of ZCS piece (a) and SEM micrographs with two magnification levels (b) and (c)



**Fig. 4** XRD of ZCS

$$D = \frac{K\lambda}{\beta \cos\theta} \dots \text{Scherrer's equation} \quad (3)$$

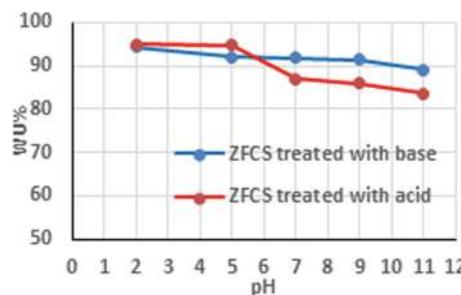
where  $D$  crystallite size(nm),  $K$  Scherrer constant (0.9),  $\lambda$  wavelength of ray(0.154 nm),  $\beta$  FWHM (radians),  $\theta$  peak position (radians), and the average crystallite size is 0.716253 nm.

### 3.5 Water Uptake Capacity (WU) OF ZCS

I submerged myself for a duration of 24 h. Results obtained from samples subjected to pretreatment with bases and acids are shown in Fig. 5. With increasing pH, the maximum water uptake percentages (WU%) for acid and base pretreatment samples

**Table 1** Data of XRD analyses of ZCS

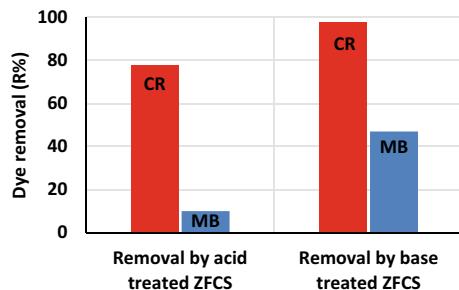
No	2θ (degree)	Intensity (count)	FWHM (2θ)	Area (count)	Integral Area (count)
1	11.02	8.0734	1.04694	7.43591	2.39316
2	12.08	20.98449	0.33732	19.15275	6.16409
3	15.74	18.19632	0.50189	13.65715	4.39539
4	16.18	14.64679	0.43676	10.57611	3.4038
5	17.56	13.78115	1.23711	13.18132	4.24226
6	21.96	22.87367	0.40125	38.02844	12.23901
7	24.62	34.63339	0.40058	48.09177	15.47778
8	28	46.09433	0.3688	30.83411	9.9236
9	28.62	32.19501	0.38011	15.42013	4.96279
10	29.98	30.11623	0.6838	36.2519	11.66726
11	33.22	16.24176	0.33706	14.70701	4.73328
12	37.44	8.50132	0.78663	17.13433	5.51449
13	38.08	9.23882	0.50251	31.2176	10.04702
14	46.88	6.72035	0.71929	15.02637	4.83606

**Fig. 5** WU% values for ZCS samples (acid and base pretreated sample

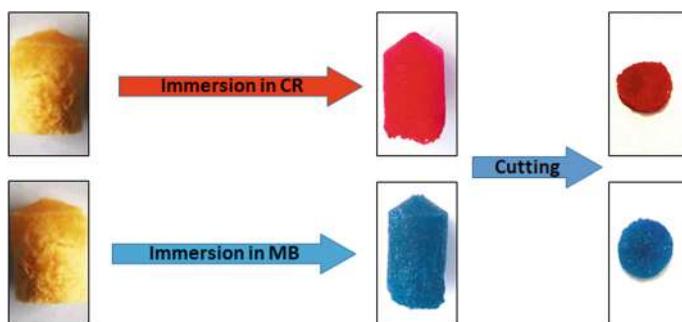
dropped from 95 and 94% to 84% and 89%, respectively. Acid-pretreated hydrogel samples exhibited a decreased water uptake percentage (WU%) in neutral and basic media compared to acidic medium. Minimal change in the proportion of bases and acids.

### 3.6 Adsorption Study

The dependence of adsorption on aqueous pH is determined by the range of 15–20. In a preliminary investigation, ZCS samples that had undergone acid and base treatment were subjected to solutions containing congo red (CR) and methylene blue (MB). To reach equilibrium, a 0.05 gramme ZCS monolith was submerged in 30 millilitres of dye solutions containing 40 parts per million (ppm)



**Fig. 6** Removal percentage values of CR and MB by ZCS (acid and base treated)



**Fig. 7** Photographs of ZCS pieces before and after immersion in dyes solutions

for 4 h. Figure 6 demonstrates that acid-treated ZCS eliminated aqueous solutions containing CR and MB at rates of 78% and 10%, respectively, whereas base treated ZCS obtained superior rates of 98% and 47%. The adsorption efficiency is greatly improved by the porous nature of zeolite. For the purpose of studying the inner surface of the ZCS and determining if dye adsorption occurs when the solution penetrates the ZCS monoliths, cylindrical samples were immersed in dye solutions, removed, and the surface water was removed using tissue paper. The specimens were sectioned to observe the cross-sectional area. Figure 7 exhibits a cross-section of the submerged monoliths, together with ZCS sections, both before to and following immersion in the dye solution. The results indicate that the dye molecules undergo diffusion into the hydrogel and then adsorb onto both its external and internal surface.

## 4 Applications

ZCS can be used in the process of filtering and purifying wastewater, because it is highly effective in removing dye.

## 5 Conclusions

ZCS is a substance that has promise for application in the treatment and purification of water in order to eliminate anionic or cationic dyes from aqueous solutions.

## References

1. Hussein WJ, Balakit AA, H.I.J.E.J.o.C. Salman (2022) Cross-linked chitosan terephthaldehyde for removal of congo red: synthesis, characterization, and adsorption studies. Egypt J Chem 65(13)
2. Salman HE, Manshad MJOGPT (2019) Methyl violet dye as corrosion inhibitor for carbon steel in acidic medium. J Glob Pharma Technol 12(01)
3. Al-Abadi SI, Al-Da'amy MA, Kareem ET (2021) Thermodynamic study for removing of crystal violet dye on Iraqi porcelanite rocks powder. In: iop conference series: earth and environmental science. IOP Publishing
4. Al-Da'amy MA, Al-Shemary RQJJGPT (2018) Removal of alizarin red dye from aqueous solution with bio sorption technique using snail shell as low cost adsorbent. J Glob Pharma Technol 10:422–430
5. Kibrahim H, Muneer A, TKreem EJJOBT (2018) Effective adsorption of azure B dye from aqueous solution using snail shell powder. J Biochem Technol 9(3):39–44
6. Crini G, Badot PMJPIPS (2008) Application of chitosan, a natural aminopolysaccharide, for dye removal from aqueous solutions by adsorption processes using batch studies: a review of recent literature. Prog Polym Sci 33(4):399–447
7. AL-Da'amy MA, AL-Khazali NA, AL-Rubaeey ETJJOGPT (2017) Removal of malachite green from aqueous solution by Iraqi porcelanite rocks. J Glob Pharma Technol 10:150–156
8. Abed A, Al Hindawi A, Alesary HJJN (2022) Green synthesis of zinc sulfide nanoparticles for the removal of methylene blue dye from aqueous solution. Nano World J 8(3):79–84
9. Shaker ZM, Salman HI, AL-Baiati MN (2023) Removal of pollutants from wastewater using a nano co-polymer surface. In: AIP conference proceedings. AIP Publishing
10. Hassanzadeh P et al (2021) Preparation and characterization of PVDF/gC 3 N 4/chitosan polymeric membrane for the removal of direct blue 14 dye. J Polym Environ 29: 3693–3702
11. Hamood ZA et al (2021) Adsorption performance of dyes over zeolite for textile wastewater treatment. Ecol Chem Eng 28(3):329–337
12. Sardar M et al (2021) Remediation of dyes from industrial wastewater using low-cost adsorbents, p 377–403
13. Mahmoodi NM, Saffar-Dastgerdi MHJM (2019) Zeolite nanoparticle as a superior adsorbent with high capacity: synthesis, surface modification and pollutant adsorption ability from wastewater. Microchem J 145:74–83
14. Athab ZH et al (2022) Enhanced macroporous cationic chitosan hydrogel by freezing and thawing method with superadsorption capacity for anionic dyes. J Polym Environ 30(9):3815–3831