

Suyel Namasudra  
Nirmalya Kar  
Sarat Kumar Patra  
David Taniar *Editors*

# Data Science and Network Engineering

Proceedings ICDSNE 2024

# **Lecture Notes in Networks and Systems**

**Volume 1165**

## **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)).

Suyel Namasudra · Nirmalya Kar ·  
Sarat Kumar Patra · David Taniar  
Editors

# Data Science and Network Engineering

Proceedings ICDSNE 2024



Springer

*Editors*

Suyel Namasudra  
Department of Computer Science  
and Engineering  
National Institute of Technology Agartala  
Agartala, Tripura, India

Sarat Kumar Patra  
Department of Electronics  
and Communication Engineering  
National Institute of Technology Rourkela  
Rourkela, Odisha, India

Nirmalya Kar   
Department of Computer Science  
and Engineering  
National Institute of Technology Agartala  
Agartala, Tripura, India

David Taniar   
Department of Information Technology  
Monash University  
Clayton, VIC, Australia

ISSN 2367-3370

Lecture Notes in Networks and Systems

ISBN 978-981-97-8335-9

<https://doi.org/10.1007/978-981-97-8336-6>

ISSN 2367-3389 (electronic)

ISBN 978-981-97-8336-6 (eBook)

© 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.

# Preface

Data science is an interdisciplinary field that uses statistics, scientific methods, scientific computing, and algorithms to extract information from potentially noisy, structured, or unstructured datasets. It is one of the fastest-growing fields because of the accelerating volume of data sources from our surroundings. Data science is popular due to its advanced features, such as decision-making, problem solving, pattern discovering, time saving, endless possibilities, cost effectiveness, and many more. It mainly combines learning techniques to solve real-life problems.

In the last two decades, there has been an advancement in Internet technologies like cloud computing, Internet of Things (IoT), big data, Virtual Reality (VR), Augmented Reality (AR), fog computing, image processing, and many more. Due to this technological development, the number of Internet users are being increased rapidly, and the amount of data generated by sensors and IoT devices is also increasing. It is very difficult to process and analyse this huge amount of data manually. This brings the need for data science techniques to automatically process and analyse data. Here, a Machine Learning (ML) algorithm helps to analyse thousands of different data points and recommend outcomes. Data science techniques can be used to monitor network traffic. Also, nowadays, many security attacks are being developed very often. It is very difficult to protect computer systems against such attacks. Here, Artificial Intelligence (AI) and ML algorithms can be used to identify attacks and safeguard systems.

This book is being planned to bring forth all the fundamental details of data science and network engineering. The main objective is to bridge the gap between the target audience (i.e., students, scientists, researchers, and academicians) from data science and network engineering domains, so that they can share ideas and recent innovations to address real-life problems. This book includes research papers presented at the Second International Conference on Data Science and Network Engineering (ICDSNE 2024) organized by the Department of Computer Science and Engineering, National Institute of Technology Agartala, Tripura, India, during July 12–13, 2024. It covers many advanced topics of data science and computer networks, such as AI, ML, Deep Learning (DL), computer networks, network performance, predictive modelling, pattern recognition, anomaly detection, blockchain, security

and privacy, Internet of Things (IoT), big data, and many more. This book includes research works of UG/PG students, researchers, academicians, business executives, and industry professionals from across the world. Many chapters of this book are presented in two different parts, which can be highly beneficial for the researchers, who are working on data science and network engineering.

Agartala, India

Agartala, India

Rourkela, India

Clayton, Australia

Suyel Namasudra

Nirmalya Kar

Sarat Kumar Patra

David Taniar

# Contents

## Computational Intelligence

<b>Hyperparameter Tuning of Fine-Tuned VGG19 Using Sine Cosine Algorithm for Acute Lymphoblastic Leukemia Detection .....</b>	<b>3</b>
Rabul Saikia, Roopam Deka, Anupam Sarma, and Salam Shuleenda Devi	
<b>Computer Vision and Deep Learning-Based Model for Detecting Spoofed Faces in Images .....</b>	<b>15</b>
Gayathri P Salian, Manasa K Rao, and M. Rashmi	
<b>Explainable AI Insights into a Time Series Weather Prediction Model Using Stacked LSTM .....</b>	<b>27</b>
T. H. Sunu Fathima and Binsu C. Kovoor	
<b>Patient-Specific and Patient-Independent Seizure Prediction Using Ensemble Learning Technique .....</b>	<b>41</b>
Ranjan Jana and Imon Mukherjee	
<b>An Intuitive and Modular Framework for Enhanced Human-Machine Synergy .....</b>	<b>53</b>
Mayank Kashyap, Apoorva Patel, Ashish Kumar, and Gurmeet Kaur	
<b>Wavelet-Transformed K-NN Pipeline for EEG-Based Eye Blink Classification with Time Wrapping .....</b>	<b>67</b>
N. Priyadarshini Jayadurga, M. Chandralekha, and Kashif Saleem	
<b>Enabling Cursor Control Through Eye Movement Using Hidden Markov Model .....</b>	<b>81</b>
G. Tanusha, P. Havarbhavi, and K. Ashwini	
<b>Comparative Study: Word2Vec Versus TF-IDF in Software Defect Predictions .....</b>	<b>95</b>
Gaurav Sharma and Priya Singh	

<b>An Improved Deep Learning Framework Based on Multi-Scale Convolutional Architecture for Road Crack Detection .....</b>	109
Idris Ya'u Idris, Badamasi Imam Ya'u, Usman Ali, and Yonis Gulzar	
<b>Knowledge Graph Relation Learning Using GAN-BERT .....</b>	123
Neelam Jain and Krupa Mehta	
<b>SightAssist: A Multi-facility Machine Learning Approach for Empowering the Visually Impaired .....</b>	133
Saikat Bandopadhyay, Jhalak Dutta, Smita Das, Soumyajit Datta, Debaditya Ghosh, Rohit Kumar Dey, and Jeet Nandigrami	
<b>Predicting COVID-19 Cases in India Using ARIMA, Prophet, LSTM and Data Analysis Using Power BI .....</b>	147
Abhrajit Das	
<b>Classification of Muti-Labeled Retinal Diseases in Retinal Fundus Images Using CNN Model ResNet18 .....</b>	163
Kowju Gayatri and Birendra Biswal	
<b>LSTM-Based Portfolio Optimization with Gerber Covariance Estimator for Increased Robustness .....</b>	179
Ishita Mehta and Kartik Gupta	
<b>Handwritten Character Recognition from Small Grayscale Images Using Pre-trained Models .....</b>	191
D. Manibharathi, C. Vasanthanayaki, and Sanjeev Kumar	
<b>Forecasting Heart Disease Using Deep BI-DI Neural Networks .....</b>	203
R. Bhuvanya, T. Kujani, and P. Matheswaran	
<b>Emo-Tune: Harnessing Emotion-Based Music for Patient Wellness .....</b>	219
T. Sudha, N. Jothy, V. Bharathi, and Santhosh Jayagopalan	
<b>Making Data Secure in Detecting ADHD with Supervised Learning .....</b>	233
Deepak Kumar Khandelwal and Mahesh Chandra Govil	
<b>An Analysis of Automatic Question Generation Research Progress and Challenges .....</b>	247
Debopam Dey and Dwijen Rudrapal	
<b>A Federated Learning Approach Towards a Privacy-Preserving Technique for Brain Tumor Classification .....</b>	259
Anurag De, Gautam Pal, Karnam Shyam, and Kalakanti Pawan Tej	
<b>Computer Networks</b>	
<b>Low Cost Emergency Communication System for Disaster Affected Areas .....</b>	275
Sanjoy Debnath, Yaddanapudi Srilekha, Vibhuthi Amarnath, and Yaddanapudi Venkata Sri Harsha	

<b>A Slot-Integrated Based Partial Ground and Tapered Patch Antenna for Satellite Communications .....</b>	<b>287</b>
J. Josiah Samuel Raj and G. Anitha	
<b>A Review on Wireless Power Transfer Systems .....</b>	<b>301</b>
Soubam Chitra Devi, Ningthoujam Juleina, Mansam Wajira, and Sorokhaibam Nilakanta Meitei	
<b>Application of Big Data to Traffic Generated in Mechanisms Containment on Optical Burst Switching Distributed Networks .....</b>	<b>313</b>
Oscar Corredor, Jannet Ortiz, Luis Ballesteros, Sergio Bermúdez, Carlos Enrique Montenegro-Marin, and Ruben Gonzalez-Crespo	
<b>A Real-Time Arm-Worn Sensor-Based Human Fall Alert Notification Model for Efficient Daily Activity Recognition .....</b>	<b>329</b>
Anurag De, C. Mugesh, Battula Lalitesh, and J. Joshua	
<b>Mobile Applications in Electronic-Healthcare: A Case Study for Bangladesh .....</b>	<b>339</b>
Jarin Tasnim, Shamim Forhad, Sunjida Mushfiq Nova, Khandakar Kamrul Hasan, S. M. Nahidul Islam, Samia Binta Hassan, Mohammad Ashiqur Noor, and Abdul Hasib Siddique	
<b>Decentralized Energy Grid System Using IoT and Blockchain: A Sustainable Future .....</b>	<b>353</b>
Saswati Debnath, Vedavati Patil, and Dhruv Agrawal	
<b>Detection and Analysis of Cyber-Attacks on IoT Network Devices .....</b>	<b>367</b>
Bashir Zak Adamu, Ilhan Firat Kilincer, and Fatih Ertam	
<b>A Hybrid Peer-to-Peer Data Center Resource Management System .....</b>	<b>383</b>
Shreyas S. Kaundinya, S. Shreyas, Joel Macklyn Dsouza, K. Gagan Prashanth, and Prafullata K. Auradkar	
<b>Author Index .....</b>	<b>397</b>

# Editors and Contributors

## About the Editors

**Suyel Namasudra** has received Ph.D. degree from the National Institute of Technology Silchar, Assam, India. He was a post-doctorate fellow at the International University of La Rioja (UNIR), Spain. Currently, Dr. Namasudra is working as an assistant professor in the Department of Computer Science and Engineering at the National Institute of Technology Agartala, Tripura, India. Before joining the National Institute of Technology Agartala, Dr. Namasudra was an assistant professor in the Department of Computer Science and Engineering at the National Institute of Technology Patna, Bihar, India. His research interests include blockchain technology, cloud computing, DNA computing, and information security. Dr. Namasudra has edited seven books, five patents, and 86 publications in conference proceedings, book chapters, and refereed journals like IEEE TII, IEEE, TNSM, IEEE TCE, IEEE T-ITS, IEEE TSC, IEEE TCSS, IEEE TCBB, ACM TOMM, ACM TOSN, ACM TALLIP, FGCS, CAEE, and many more.

**Nirmalya Kar** a member of IEEE, is currently an Assistant Professor in the Department of Computer Science and Engineering at the National Institute of Technology (NIT) Agartala and designated Chief Information Security Officer. Having more than 18 years of teaching and research experience, he specializes in Information Security, Cryptography, DNA Computing, Computational Intelligence and the Internet of Things (IoT). He has 2 patents and more than 70+ research contributions in book chapters, conference proceedings, and refereed journals like IEEE TNSM, IEEE TCE, ACM TOMM, and many more. Dr. Kar has been involved in various academic roles including Organising Chair and General Chair of International Conferences, Editor of Book Chapters, reviewer board of many SCI journals and transactions apart from other administrative roles like coordinating high-performance computing initiatives at NIT Agartala, Academic coordinator of UG curriculum, etc.

**Sarat Kumar Patra** received Ph.D. degree from the Edinburgh University, UK, in 1998. He is a professor in the Department of Electronics and Communication Engineering at the NIT Rourkela, India. Professor Patra has a total of 37 years of experience in teaching and industry. His research interests include wireless communication, soft computing, optical communication, and deep learning. Professor Patra has edited many books and more than 200 publications in conference proceedings, book chapters, and refereed journals. His h-index is 20. Currently, he holds the position of Director at the National Institute of Technology Agartala from May 2023. He also served as the director of IIIT Vadodara from July 2017 to May 2023.

**David Taniar** received all his degrees (Bachelor, Master, and Ph.D.) in Computer Science. His research expertise includes data warehousing, data management, data engineering, and data analytics. His recent book on Data Warehousing and Analytics (Springer, 2021) has been accessed more than 50 thousand times, and it is being used as a textbook worldwide. He has graduated more than 25 Ph.D. students in his career. He is currently an associate professor at Monash University, Australia.

## Contributors

**Bashir Zak Adamu** Department of Digital Forensics Engineering, Firat University, Elazig, Türkiye

**Dhruv Agrawal** Department of Computer Science and Engineering, Alliance University, Bengaluru, Karnataka, India

**Usman Ali** Department of Computer Science, School of Science, Federal College of Education (Technical), Gombe, Nigeria

**Vibhuthi Amarnath** Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India

**G. Anitha** Centre for Applied Research, Institute of Electronics and Communication Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Chennai, Tamil Nadu, India

**K. Ashwini** Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Chennai, India

**Prafullata K. Auradkar** Department of CSE, PES University, Bengaluru, India

**Luis Ballesteros** Universidad Cooperativa de Colombia, Bogotá, Colombia

**Saikat Bandopadhyay** Department of Artificial Intelligence and Machine Learning, Netaji Subhash Engineering College, Kolkata, India

**Sergio Bermúdez** Universidad Cooperativa de Colombia, Bogotá, Colombia

**V. Bharathi** Sri Manakula Vinayagar Engineering College, Puducherry, India

**R. Bhuvanya** Sri Ramachandra Faculty of Engineering and Technology, Sri Ramachandra Institute of Higher Education and Research, Chennai, Tamil Nadu, India

**Birendra Biswal** Centre for Medical Imaging Studies, Department of ECE, Gayatri Vidya Parishad College of Engineering (A), Visakhapatnam, India

**M. Chandrakha** Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Chennai, India

**Oscar Corredor** Universidad Cooperativa de Colombia, Bogotá, Colombia

**Abhrajit Das** Seidenberg School of Computer Science and Information Systems, Pace University, New York, USA

**Smita Das** Department of Computer Science and Engineering, National Institute of Technology, Agartala, India

**Soumyajit Datta** Department of Artificial Intelligence and Machine Learning, Heritage Institute of Technology, Kolkata, India

**Anurag De** School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh, India

**Sanjoy Debnath** Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India

**Saswati Debnath** Department of Computer Science and Engineering, Alliance University, Bengaluru, Karnataka, India

**Roopam Deka** Department of Pathology and Lab Medicine, All India Institute of Medical Sciences, Guwahati, India

**Salam Shuleenda Devi** Department of Electronics and Communication Engineering, National Institute of Technology Meghalaya, Shillong, Meghalaya, India

**Soubam Chitra Devi** Department of Electrical Engineering, Manipur Institute of Technology, Manipur University, Canchipur, Imphal, India

**Debopam Dey** Department of CSE, NIT Agartala, Agartala, India

**Rohit Kumar Dey** Department of Computer Science and Engineering, Heritage Institute of Technology, Kolkata, India

**Jhalak Dutta** Department of Computer Science and Engineering, Heritage Institute of Technology, Kolkata, India

**Fatih Ertam** Department of Digital Forensics Engineering, Firat University, Elazig, Türkiye

**Shamim Forhad** University of Scholars, Dhaka, Bangladesh

**K. Gagan Prashanth** Department of CSE, PES University, Bengaluru, India

**Kowju Gayatri** Andhra University, Visakhapatnam, India;  
Centre for Medical Imaging Studies, Department of ECE, Gayatri Vidya Parishad  
College of Engineering (A), Visakhapatnam, India

**Debaditya Ghosh** Department of Artificial Intelligence and Machine Learning,  
Heritage Institute of Technology, Kolkata, India

**Ruben Gonzalez-Crespo** Universidad Internacional de La Rioja, Madrid, España

**Mahesh Chandra Govil** National Institute of Technology, Sikkim, Ravangla,  
India

**Yonis Gulzar** Department of Management Information Systems, College of  
Business Administration, King Faisal University, Al Ahsa, Saudi Arabia

**Kartik Gupta** Department of Software Engineering, Delhi Technological  
University, New Delhi, Delhi, India

**Yaddanapudi Venkata Sri Harsha** Vel Tech Rangarajan Dr. Sagunthala R&D  
Institute of Science and Technology, Chennai, India

**Khandakar Kamrul Hasan** University of Scholars, Dhaka, Bangladesh

**Samia Binta Hassan** University of Scholars, Dhaka, Bangladesh

**P. Havarbhavi** Department of Computer Science and Engineering, Amrita School  
of Computing, Amrita Vishwa Vidyapeetham, Chennai, India

**Idris Ya'u Idris** Department of Computer Science, School of Science and  
Technology, Federal Polytechnic Bauchi, Bauchi, Nigeria

**S. M. Nahidul Islam** University of Scholars, Dhaka, Bangladesh

**Neelam Jain** SVKM's Mithibai College of Arts, Chauhan Institute of Science &  
Amrutben Jivanlal College of Commerce And Economics, Mumbai, India

**Ranjan Jana** Indian Institute of Information Technology, Kalyani, India;  
RCC Institute of Information Technology, Kolkata, India

**Santhosh Jayagopalan** British Applied College, Umm Al Quwain, UAE

**J. Joshua** School of Computer Science and Engineering, VIT-AP University,  
Amaravati, Andhra Pradesh, India

**J. Josiah Samuel Raj** Centre for Applied Research, Institute of Electronics and  
Communication Engineering, Saveetha School of Engineering, Saveetha Institute  
of Medical and Technical Sciences (SIMATS), Saveetha University, Chennai,  
Tamil Nadu, India

**N. Jothy** SRM Valliammai Engineering College, Kattangulathur, Tamil Nadu,  
India

**Ningthoujam Juleina** Department of Electrical Engineering, Manipur Institute of Technology, Manipur University, Canchipur, Imphal, India

**Mayank Kashyap** Department of Computer Science and Engineering, Chandigarh University, Mohali, Punjab, India

**Shreyas S. Kaundinya** Department of CSE, PES University, Bengaluru, India

**Gurmeet Kaur** Department of Computer Science and Engineering, Chandigarh University, Mohali, Punjab, India

**Deepak Kumar Khandelwal** National Institute of Technology, Sikkim, Ravangla, India

**Ilhan Firat Kilincer** Department of Digital Forensics Engineering, Firat University, Elazig, Türkiye

**Binsu C. Kovoor** Division of Information Technology, Cochin University of Science and Technology, Kochi, Kerala, India

**T. Kujani** VelTech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India

**Ashish Kumar** Department of Computer Science and Engineering, Chandigarh University, Mohali, Punjab, India

**Sanjeev Kumar** University of Illinois, Urbana-Champaign, USA

**Battula Lalitesh** School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh, India

**Joel Macklyn Dsouza** Department of CSE, PES University, Bengaluru, India

**D. Manibharathi** Government College of Engineering, Salem, India

**P. Matheswaran** K. Ramakrishnan College of Technology, Samayapuram, Tiruchirappalli, Tamil Nadu, India

**Ishita Mehta** Department of Software Engineering, Delhi Technological University, New Delhi, Delhi, India

**Krupa Mehta** Faculty of Computer Application and IT, GLS University, Ahmedabad, Gujarat, India

**Sorokhaibam Nilakanta Meitei** Department of Electrical Engineering, Manipur Institute of Technology, Manipur University, Canchipur, Imphal, India

**Carlos Enrique Montenegro-Marin** Universidad Distrital Francisco José de Caldas, Bogotá, Colombia

**C. Mugesha** School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh, India

**Imon Mukherjee** Indian Institute of Information Technology, Kalyani, India

**Jeet Nandigrami** Department of Artificial Intelligence and Machine Learning, Heritage Institute of Technology, Kolkata, India

**Mohammad Ashiqur Noor** University of Scholars, Dhaka, Bangladesh

**Sunjida Mushfiq Nova** University of Scholars, Dhaka, Bangladesh

**Jannet Ortiz** Universidad Cooperativa de Colombia, Bogotá, Colombia

**Gautam Pal** Computer Science and Engineering Department, TIT Narsingarh, Agartala, Tripura, India

**Apoorva Patel** Department of Computer Science and Engineering, Chandigarh University, Mohali, Punjab, India

**Vedavati Patil** Department of Computer Science and Engineering, Alliance University, Bengaluru, Karnataka, India

**N. Priyadarshini Jayadurga** Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Chennai, India

**Manasa K Rao** Department of Data Science and Computer Applications, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India

**M. Rashmi** Department of Data Science and Computer Applications, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India

**Dwijen Rudrapal** Department of CSE, NIT Agartala, Agartala, India

**Rabul Saikia** Department of Electronics and Communication Engineering, National Institute of Technology Meghalaya, Shillong, Meghalaya, India

**Kashif Saleem** Department of Computer Sciences and Engineering, College of Applied Studies and Community Service, King Saud University, Riyadh, Saudi Arabia

**Gayathri P Salian** Department of Data Science and Computer Applications, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India

**Anupam Sarma** Department of Onco-Pathology, Dr. Bhubaneswar Borooah Cancer Institute, Guwahati, India

**Gaurav Sharma** Delhi Technological University, Delhi, India

**S. Shreyas** Department of CSE, PES University, Bengaluru, India

**Karnam Shyam** School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh, India

**Abdul Hasib Siddique** University of Scholars, Dhaka, Bangladesh

**Priya Singh** Delhi Technological University, Delhi, India

**Yaddanapudi Srilekha** Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India

**T. Sudha** Sri Manakula Vinayagar Engineering College, Puducherry, India

**T. H. Sunu Fathima** Division of Information Technology, Cochin University of Science and Technology, Kochi, Kerala, India;  
Department of Computer Science, Rajagiri College of Social Sciences (Autonomous), Kalamassery, Kerala, India

**G. Tanusha** Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Chennai, India

**Jarin Tasnim** University of Scholars, Dhaka, Bangladesh

**Kalakanti Pawan Tej** School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh, India

**C. Vasanthanayaki** Government College of Engineering, Bodinayakkanur, India

**Mansam Wajira** Department of Electrical Engineering, Manipur Institute of Technology, Manipur University, Cachhipur, Imphal, India

**Badamasi Imam Ya'u** Department of Mathematical Sciences, Faculty of Science, Abubakar Tafawa Balewa University, Bauchi, Nigeria

# **Computational Intelligence**

# Hyperparameter Tuning of Fine-Tuned VGG19 Using Sine Cosine Algorithm for Acute Lymphoblastic Leukemia Detection



Rabul Saikia, Roopam Deka, Anupam Sarma, and Salam Shuleenda Devi

**Abstract** Acute Lymphoblastic Leukemia (ALL) is a highly malignant disorder that explicitly affects White Blood Cells (WBC) produced from lymphoid cells. Due to its potentially fatal nature, the prompt diagnosis of ALL is of utmost importance. At present, the domain of medical sciences is profoundly influenced by automated computer-assisted methodologies that are based on Artificial Intelligence (AI) and Deep Learning (DL). These methodologies serve as indispensable instruments for physicians in the expeditious identification of illnesses and also in reducing their burden. This paper proposed a transfer learning-based methodology where the last 4 convolution layers of VGG19 are fine-tuned. Next, the best hyperparameters of the fine-tuned VGG19 are evaluated using the Sine Cosine Algorithm (SCA). Further, a mixed dataset is proposed by combining a private dataset comprising 1000 smear images with a public dataset comprising 108 images. The overall scheme was experimented on the mixed dataset that achieved an accuracy of 98.18%. Moreover, the comparative studies establish the superiority of the proposed ALL scheme over comparing methods.

**Keywords** Acute Lymphoblastic Leukemia (ALL) · VGG19 · Fine-Tuning · Sine Cosine Algorithm (SCA)

---

R. Saikia · S. S. Devi (✉)

Department of Electronics and Communication Engineering, National Institute of Technology

Meghalaya, Shillong, Meghalaya 793003, India

e-mail: [sshuleendadevi@nitm.ac.in](mailto:sshuleendadevi@nitm.ac.in)

R. Saikia

e-mail: [p21ec003@nitm.ac.in](mailto:p21ec003@nitm.ac.in)

R. Deka

Department of Pathology and Lab Medicine, All India Institute of Medical Sciences,  
Guwahati 781101, India

e-mail: [roopamdeka@aiimsguwahati.ac.in](mailto:roopamdeka@aiimsguwahati.ac.in)

A. Sarma

Department of Onco-Pathology, Dr. Bhubaneswar Borooh Cancer Institute, Guwahati 781016,  
India

e-mail: [dranupamsarma@gmail.com](mailto:dranupamsarma@gmail.com)

## 1 Introduction

Leukemia is characterized as a potentially fatal illness that disrupts the regular operation of tissues responsible for producing blood. The majority of WBC in the body becomes malignant in leukemia. WBC may be categorized into two main types specifically lymphoid as well as myeloid cells. Lymphoblastic leukemia occurs if lymphoid cells undergo malignance, whereas myeloid leukemia occurs when myeloid cells become malignant. The progression rate of malignant cells may fluctuate, and, based on this criterion, leukemia may be further classified as acute and chronic type. In acute leukemia, the progression rate of leukemic cells is high, whereas in chronic leukemia, the progression rate of leukemic cells is moderate. Thus, there are four main kinds of leukemia: acute myeloid leukemia (AML), acute lymphoblastic leukemia (ALL), chronic myeloid leukemia (CML), and chronic lymphocytic leukemia (CLL). Leukemia has the potential to impact individuals of all age groups. However, it is often identified in individuals under the age of 15 and adults over the age of 55. The symptoms exhibited by a leukemia patient include fever, weakness, and fatigue, which are similar to the symptoms exhibited by a patient with the flu [1]. Consequently, this complicates the process of diagnosing leukemia during its first phase.

Annually, India reports around 10,000 new cases of pediatric leukemia on average. According to the GLOBOCAN 2018 report, almost 0.3 lakh individuals worldwide succumbed to leukemia [2]. Timely identification of the illness may prevent a patient from succumbing to a life-threatening infection, which might result in untimely demise [3]. The use of machine learning (ML) and deep learning (DL)-based automated computer-assisted diagnostic systems (CAS) assists pathologists in identifying and categorizing diseases in the early stages and reducing their workload [4, 5]. Thus, further developmental work on the CAS is of utmost importance to assist pathologists in early illness detection with the most reliable results. This is the motivation behind this work.

The ML-based leukemia detection approaches [6–8] depend on crucial stages like segmentation, user-defined feature computation, and classification to attain significant results. In leukemia detection, WBC segmentation is a challenging task due to inhomogeneity among blood cells. On the other hand, the DL methods outperform classical ML methods for the diagnosis of leukemia without using segmentation. However, DL approaches in the biomedical field are the paucity of large, adequately annotated datasets [9, 13]. Moreover, the proper tuning of hyperparameters (learning rate, batch size, maximum epoch, and momentum) is of utmost importance to the DL-based model for better convergence and to achieve optimum performance. To address these challenges, SCA-based fine-tuned VGG19 is proposed with the following contributions:

- **Mixed Dataset:** It is a proposed mixed dataset by combining a private dataset with an ALL-IDB1 [16] public dataset. The private dataset is collected from Dr. Bhubaneswar Borooah Cancer Institute (Dr. BBCI), Guwahati, India, comprising 1000 images. The ALL-IDB1 dataset comprises 108 images.

- Proposed a pretrained VGG19-based ALL detection model by fine-tuning its last 4 convolution layers and FC layers.
- Metaheuristic optimization strategy, the Sine Cosine Algorithm (SCA) is employed for hyperparameter tuning of the proposed ALL detection scheme.
- Comparative analysis is done with three perspectives: initially, the performance of the proposed methodology with other optimization algorithms; secondly, the performance of the proposed methodology with other pretrained CNNs; and finally, the proposed scheme is compared with the existing leukemia detection methods in the literature.

The rest of the paper is structured like this: Sect. 2 illustrates related works; Sect. 3 illustrates data acquisition and the adapted proposed methodology; Sect. 4 illustrates experimental outcomes with comparison analysis; and Sect. 5 illustrates the conclusion with future scopes.

## 2 Related Work

Leukemia image analysis is a growing area that has attracted a lot of interest and resources recently. Mohapatra et al. [6] utilized single WBC images by considering the combination of features like morphological, Haar wavelet, Haralick's, 2D-Discrete Fourier Transform (2D-DFT), and color features in conjunction with an ensemble classifier. The accuracy of their model was reported to be 94.73%. Putzu et al. [7] used thresholding-based techniques, namely Zack and Otsu algorithms, for the segmentation of WBC. A total of 131 features, encompassing shape, color, and texture descriptors, were retrieved and used to identify ALL and healthy cases using a Support Vector Machine (SVM) classifier with an RBF kernel. The classification was performed using the ALL-IDB1 public dataset, resulting in an accuracy of 93.20%. Patel and Mishra [8] detected WBCs using K-mean clustering and applied histogram equalization with the Zack method for grouped WBC. The features, specifically mean, standard deviation, color, area, and perimeter, were computed with SVM classification, obtaining a 93.57% accuracy.

There have been recent works showing that DL approaches can automate feature extraction [5, 9]. In order to distinguish between healthy and acute leukemia patients, Abhishek et al. [9] used pretrained DenseNet121 with SVM that achieved a 98.00% accuracy rate using a dataset consisting of 608 microscopic images. Anilkumar et al. [10] developed a custom CNN called “LeukNet” that is based on the architecture of AlexNet. They used the ASH dataset to identify B-ALL and T-ALL cells and achieved an accuracy of 94.12%. Shafique and Tehsin [11] used a pretrained AlexNet approach that was adjusted and a 96.06% classification accuracy on the ALL-IDB 2 dataset was obtained. Thanh et al. [12] used a private dataset consisting of 1188 smear images to develop a CNN approach for distinguishing between healthy and malignant cells. Their suggested technique achieved a 96.60% accuracy rate. Shahin et al. [13] suggested a custom CNN strategy called WBCsNet, which utilizes transfer

learning. The architecture consists of 3 convolution layers, 2 pooling levels, 4 activation layers (ReLU), and 2 Fully Connected (FC) layers. A SoftMax layer was then used to classify ALL and healthy WBC. The approach was implemented on both the ALL-IDB and private datasets, resulting in an accuracy rate of 96.10%. Das and Meher [14] employed the transfer-learning model ResNet50 to extract automated features followed by the classification using Logistic Regression (LR), SVM, and Random Forest (RF). The accuracy achieved for all classes was 96.15% on the ALL-IDB2 dataset. In another work, Das and Meher [15] created an ALL-detection system using a hybrid transfer learning method that combines two CNN architectures, MobileNetV2 and ResNet18. The hybrid model exhibited a remarkable degree of performance on the ALLIDB2 dataset, with an accuracy of 97.18% and a recall of 95.90%. Das and Meher [15] suggested deep spatial features adapting ResNet-101, GoogleNet, SqueezeNet, DenseNet-201, and MobileNetV2. Next, the temporal features from the Gated Recurrent Unit (GRU) and Bidirectional Long Short-Term Memory (Bi-LSTM) are combined with these features. The DenseNet-201 with Multiclass SVM (MSVM) provided an F1-score of 96.23% and an accuracy of 96.29% on the C\_NMC\_2019 dataset. Mohammed et al. [16] proposed an ALL detection strategy using a residual network in the first block of CapsNet. The model has 96.97% specificity, 96.81% sensitivity, 96.79% precision, and 97.44% accuracy across all three datasets (ALL-IDB1, ALL-IDB2, C\_NMC\_2019) after 20% pruning.

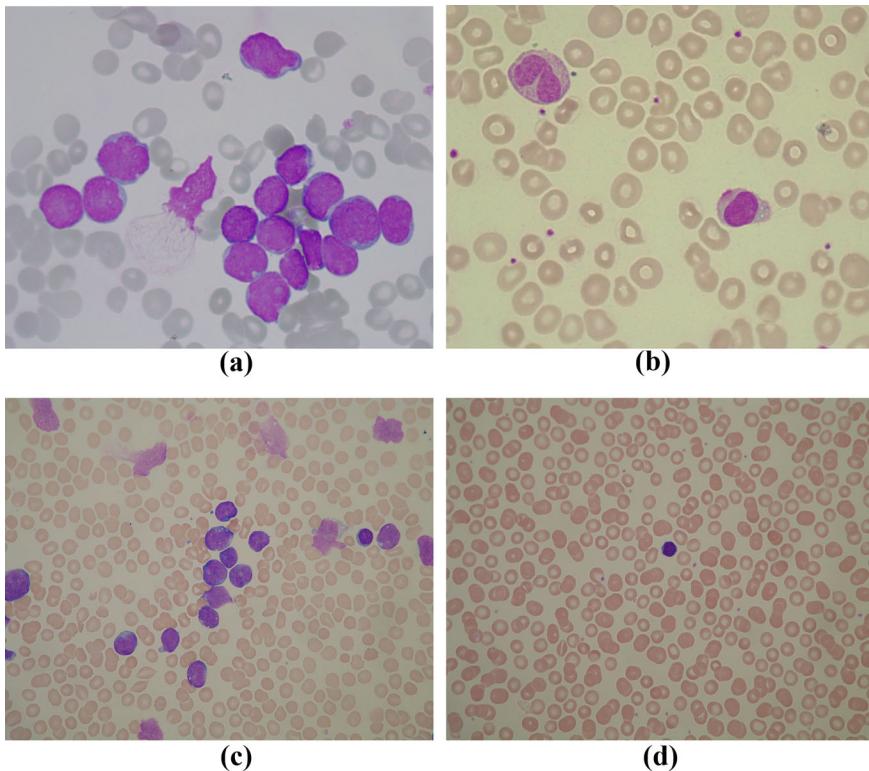
### 3 Methodology

#### 3.1 Data Acquisition (*Mixed Dataset*)

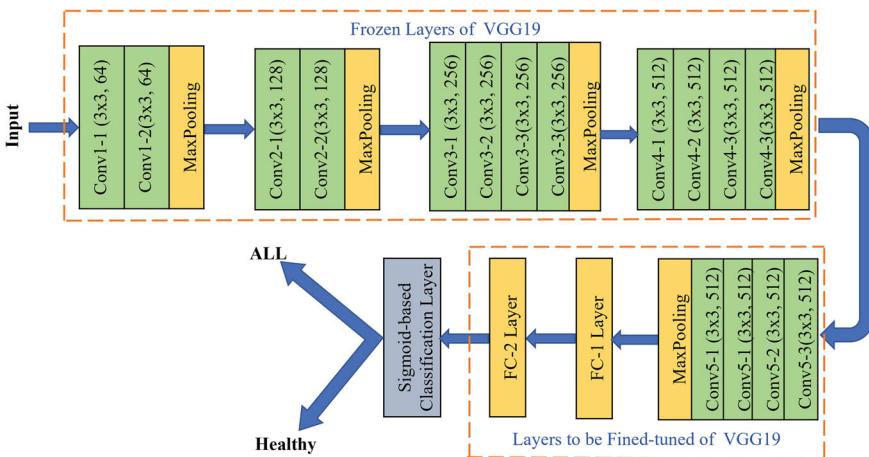
The dataset utilized for the proposed methodology is created by combining two datasets. The first dataset is collected from Dr. Bhubaneswar Borooah Cancer Institute (Dr BBCI), Guwahati, India, with prior approval of the Institutional Ethics Committee (IEC). This dataset comprises 500 images belonging to the ALL category and 500 images belonging to the healthy category. The second dataset is the ALL-IDB1 [18] public dataset, which comprises 49 images belonging to the ALL category and 59 images belonging to the healthy category. The sample of smear images is depicted in Fig. 1.

#### 3.2 Proposed Methodology

The proposed ALL detection scheme is constructed based on VGG19 [19], where the last 4 convolution layers along with FC layers are fine-tuned. The overall scheme is displayed in Fig. 2. Further, its hyperparameters are tuned using SCA as illustrated in Sect. 3.2.1.



**Fig. 1** Sample of smear images from Dr. BBCI (a–b) and ALL-IDB1 (c–d). The first column represents the ALL category, and the second column represents the Healthy category



**Fig. 2** Proposed scheme to detect ALL

### 3.2.1 Sine Cosine Algorithm (SCA)

SCA [20] is a metaheuristic optimization technique that uses trigonometric functions, Sine and Cosine, to generate many initial random candidate solutions and guides them to oscillate away from or toward the optimal answer. It employs the following equations to update the solutions in the search domain:

$$Z_{jk}^{p+1} = Z_{jk}^p + \rho_1 \sin(\rho_2) \left| \rho_3 B_{jk}^p - Z_{jk}^p \right| \quad (1)$$

$$Z_{jk}^{p+1} = Z_{jk}^p + \rho_1 \cos(\rho_2) \left| \rho_3 B_{jk}^p - Z_{jk}^p \right| \quad (2)$$

Equations (1) and (2) can be combined as follows:

$$Z_{jk}^{p+1} = \begin{cases} Z_{jk}^p + \rho_1 \cos(\rho_2) \left| \rho_3 B_{jk}^p - Z_{jk}^p \right|, & \rho_4 \geq 0.5 \\ Z_{jk}^p + \rho_1 \sin(\rho_2) \left| \rho_3 B_{jk}^p - Z_{jk}^p \right|, & \rho_4 < 0.5 \end{cases} \quad (3)$$

where  $Z_{jk}^p$  denotes current individual  $j$  at iteration  $p$  and  $B_{jk}^p$  denotes best individual position at iteration  $p$ . The symbols  $\rho_1$ ,  $\rho_2$ ,  $\rho_3$ , and  $\rho_4$  represent the random parameters.

The value of  $\rho_1$  determines whether the solution is updated toward the optimal solution  $\rho_1 < 1$  or away from it  $\rho_1 > 1$ . The variable  $\rho_1$  exhibits a linear drop from a certain constant value to zero:

$$\rho_1 = c - p \frac{c}{P_{max}} \quad (4)$$

In Eq. (4),  $c$  represents a constant term,  $p$  represents the current iteration, and  $P_{max}$  represents the maximum allowed iteration.

The variable  $\rho_2$  takes on values between 0 and  $2\pi$ , and it controls the magnitude of the movement of a solution toward or away from the goal. The  $\rho_3$  value ranges from 0 to 2 and assigns weight to the destination point, controlling its influence on the updating process of other solutions. Finally, with values ranging from 0 to 1,  $\rho_4$  switches between sine and cosine functions. The initial parameters of SCA for the proposed methodology are chosen as displayed in Table 1.

## 4 Results and Discussion

The proposed ALL detection system starts by splitting the mixed dataset into training, validation, and test sets by randomly dividing it into three distinct sets, with 70% for training, 10% for validation, and 20% for testing. Table 2 summarizes the data partitioning of the mixed dataset. Min–max scaling [21] is employed to normalize

**Table 1** Initial parameters of SCA

Parameters	Value
$c$	2 to 0
$P_{max}$	20
$\rho_2$	$[0, 2\pi]$
$\rho_3$	0 to 2
$\rho_4$	0 to 1
Population size	10

**Table 2** Splitting summary of the mixed dataset

Classes	Train valid test		
ALL	385	55	109
Healthy	392	56	111
Total	777	111	220

the images since the images are acquired at various time durations, various lighting conditions, and various camera specifications. Further, the images are incorporated with augmentation techniques to handle the overfitting issue.

The experiments are carried out using Python-based Keras [22] toolkit in a workstation with a 3.20 GHz Intel Xeon processor, 32 GB of RAM, and a 16 GB dedicated GPU memory. The performance metrics considered for the proposed method are illustrated in Table 3, where  $TR_P$  denotes correct classification as positive class,  $TR_N$  denotes correct classification as negative class,  $FA_P$  denotes incorrect classification as positive class, and  $FA_N$  denotes incorrect classification as negative class. The proposed methodology is a binary classification task; thus, Sigmoid function is used in the output layer. The images are resized to  $224 \times 224$  pixels, and the Stochastic Gradient Descent (SGD) [23] is used as an optimizer with the binary cross-entropy loss function. Next, the SCA method is applied for tuning the hyperparameters, namely batch size, epochs, momentum, and learning rate. The values of hyperparameters obtained through the SCA algorithm are illustrated in Table 4, which provides the best results.

Table 5 displays the performance of the proposed SCA-based ALL-classification model along with other optimization algorithms like Particle Swarm Optimization

**Table 3** Performance metrics employed in this work

Parameters	Formulae
Accuracy	$\frac{TR_P + TR_N}{TR_P + TR_N + FA_P + FA_N}$
Precision	$\frac{TR_P}{TR_P + FA_P}$
Recall	$\frac{TR_P}{TR_P + FA_N}$
F1-Score	$\frac{2 \times Precision \times Recall}{Precision + Recall}$
Specificity	$\frac{TR_N}{TR_N + FA_P}$

**Table 4** Details of hyperparameters obtained through the SCA

Optimization technique	Initial learning rate	Batch size	Maximum epoch	Momentum
SCA	0.00127	16	22	0.79

(PSO) [24] and Gray Wolf Optimizer (GWO) [25]. By analyzing Table 5, it becomes evident that the proposed methodology achieved superior performance with a 98.18% accuracy rate. Further, it achieved excellent performance in terms of precision, recall, specificity, and F1-score. The experimental outcomes signify that the proposed method attained accuracy enhancement of 1.82 and 1.36% compared to the other optimization strategies, PSO and GWO, respectively.

Table 6 represents the confusion matrix of the proposed method and other optimization strategies using  $TR_N$ ,  $FA_P$ ,  $FA_N$ , and  $TR_P$  values. By analyzing Table 6, it signifies that the proposed methodology successfully reduces the number of false detections.

The corresponding Receiver Operating Characteristic (ROC) curves for the proposed method and other optimization techniques are displayed in Fig. 3. Classifier performance is evaluated by analyzing the area contained by the ROC curve and the discrete diagonal line, which represents a random classifier. One numerical measure of classifier performance is the area under the ROC curve (AUC). A random classifier performs 0.5, whereas a perfect classifier performs 1.0. Figure 3 signifies that the proposed methodology attains superior performance with an AUC value of 0.981 compared to other optimization strategies.

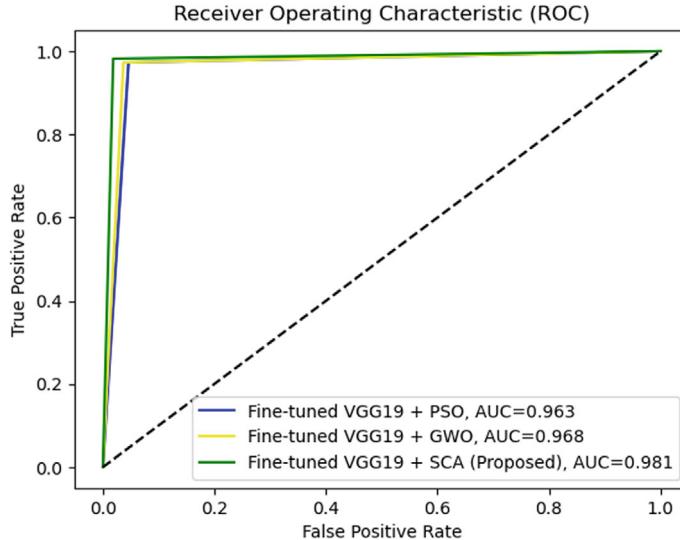
The performance of the proposed scheme is compared to some pretrained CNNs like ResNet50 [26], DenseNet121 [27], and MobileNet [28] as demonstrated in

**Table 5** Performance of the proposed scheme and other optimization strategies

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Specificity (%)
Fine-tuned VGG19 + PSO	96.36	95.58	97.30	96.43	95.41
Fine-tuned VGG19 + GWO	96.82	96.43	97.30	96.86	96.33
<b>Fine-tuned VGG19 + SCA (Proposed)</b>	<b>98.18</b>	<b>98.20</b>	<b>98.20</b>	<b>98.20</b>	<b>98.17</b>

**Table 6** Performance of the proposed method and other optimization algorithms in terms of  $TR_N$ ,  $FA_P$ ,  $FA_N$ , and  $TR_P$ 

Methods	$TR_N$	$FA_P$	$FA_N$	$TR_P$
Fine-tuned VGG19 + PSO	104	5	3	108
Fine-tuned VGG19 + GWO	105	4	3	108
<b>Fine-tuned VGG19 + SCA (Proposed)</b>	<b>107</b>	<b>2</b>	<b>2</b>	<b>109</b>



**Fig. 3** ROC curves of the proposed scheme and other optimization techniques

Table 7. It signifies that the proposed methodology attained superior performance compared to ResNet50, DenseNet121, and MobileNet with accuracy improvements of 10.00, 2.73, and 2.27%, respectively.

Table 8 represents the confusion matrix of the proposed method and other pretrained CNNs using  $TR_N$ ,  $FA_P$ ,  $FA_N$ , and  $TR_P$  values. By analyzing Table 8, it signifies that the proposed methodology successfully reduces the number of false detections.

**Table 7** Performance comparison of the proposed scheme with pretrained CNNs

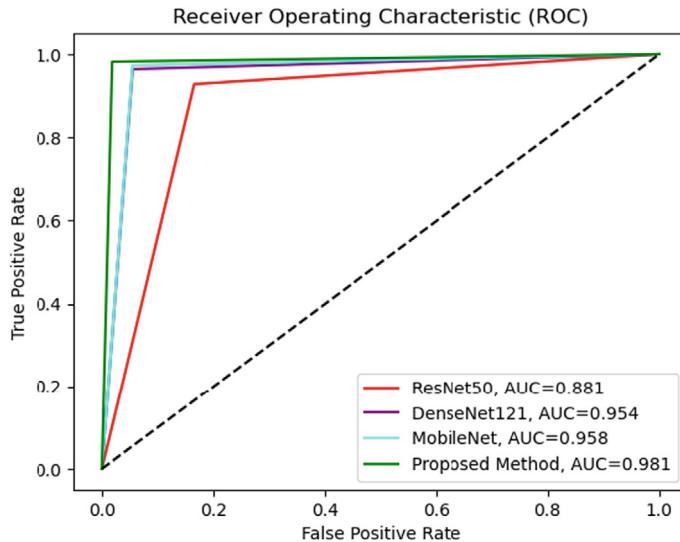
Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Specificity (%)
ResNet50	88.18	85.12	92.79	88.79	83.49
DenseNet121	95.45	94.69	96.40	95.54	94.50
MobileNet	95.91	94.74	97.30	96.00	94.50
<b>Proposed</b>	<b>98.18</b>	<b>98.20</b>	<b>98.20</b>	<b>98.20</b>	<b>98.17</b>

**Table 8** Performance comparison of the proposed method and other pretrained CNNs in terms of  $TR_N$ ,  $FA_P$ ,  $FA_N$ , and  $TR_P$

Methods	$TR_N$	$FA_P$	$FA_N$	$TR_P$
ResNet50	91	18	8	103
DenseNet121	103	6	4	107
MobileNet	103	6	3	108
<b>Proposed</b>	107	2	2	109

The ROC curves for the proposed method and other pretrained CNNs are displayed in Fig. 4. It signifies that proposed methodology attains superior performance with AUC value 0.981.

The performance evaluation of proposed ALL detection scheme is compared with the existing leukemia detection methodologies available in the literature as demonstrated in Table 9. From Table 9, it is established that the proposed ALL detection scheme outperforms the existing methodologies with improvements in accuracy 4.06, 2.12, 2.08, 2.03, 1.89, and 0.74% compared to [10, 11, 13, 14, 16, 17], respectively.



**Fig. 4** ROC curves of the proposed scheme and other pretrained CNNs

**Table 9** Accuracy comparison of the proposed scheme with existing leukemia detection methods

Authors & Years	Approaches	Used dataset	Accuracy (%)
Anilkumar et al. [10]	LeukNet	ASH	94.12
Shafique and Tehsin [11]	AlexNet	ALL-IDB2	96.06
Shahin et al. [13]	WBCsNet	ALL-IDB	96.10
Das and Meher [14]	ResNet50	ALL-IDB2	96.15
Mohammed et al. [16]	DenseNet201 + MSVM	C_NMC_2019	96.29
Dhalla et al. [17]	CapsNet	ALL-IDB1, ALL-IDB2, C_NMC_2019	97.44
<b>Proposed</b>	<b>SCA-based fine-tuned VGG19</b>	<b>Mixed Dataset (1108 images)</b>	<b>98.18</b>

## 5 Conclusion

The proposed work in this paper presents an ALL-detection methodology using fine-tuned VGG19 along with the SCA strategy for finding its optimal hyperparameters. The proposed technique used a mixed dataset generated by combining two datasets: one sourced from Dr. BBCI Guwahati and the other from the public dataset ALL-IDB1. The performance of the proposed scheme is evaluated using performance metrics like accuracy, precision, recall, F1-score, and specificity. The experimental outcomes signify that the proposed ALL-detection strategy obtained excellent performance with a reported accuracy of 98.18%. The comparison analysis demonstrates that the proposed methodology surpassed the optimizing approaches like PSO and GWO. Moreover, it demonstrated superior performance in comparison to pretrained CNN like ResNet50, DenseNet121, and MobileNet. In the future, the proposed scheme might be expanded to a multi-class classification framework by including another acute leukemia class, namely AML.

**Acknowledgements** The authors would like to acknowledge Dr. Bhubaneswar Borooh Cancer Institute (Dr. BBCI), Guwahati, India, for their valuable assistance in gathering the dataset under the Institutional Ethics Committee (IEC) No. “Ref. No. BBCI-TMC/Misc-01/MEC/125/2022”.

## References

1. Agaian S, Madhukar M, Chronopoulos AT (2014) Automated screening system for acute myelogenous leukemia detection in blood microscopic images. IEEE Syst J 8(3):995–1004
2. Cancer Today, <https://gco.iarc.fr/today/data/factsheets/populations/356-india-fact-sheets.pdf>. Accessed 7 Nov 2023
3. Vogado LH, Veras RM, Araujo FH, Silva RR, Aires KR (2018) Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification. Eng Appl Artif Intell 72:415–422
4. Abhishek A, Jha RK, Sinha R, Jha K (2023) Automated detection and classification of leukemia on a subject-independent test dataset using deep transfer learning supported by Grad-CAM visualization. Biomed Signal Process Control 83:104722
5. Deb SD, Jha RK (2020) Covid-19 detection from chest x-ray images using ensemble of CNN models. In: International conference on power, instrumentation, control and computing (PICC). IEEE, Thrissur, India, pp 1–5
6. Mohapatra S, Patra D, Satpathy S (2014) An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images. Neural Comput Appl 24(7):1887–1904
7. Putzu L, Caocci G, Di Ruberto C (2014) Leucocyte classification for leukaemia detection using image processing techniques. Artif Intell Med 62(3)
8. Patel N, Mishra A (2015) Automated leukaemia detection using microscopic images. Procedia Comput Sci 58:635–642
9. Abhishek A, Jha RK, Sinha R, Jha K (2022) Automated classification of acute leukemia on a heterogeneous dataset using machine learning and deep learning techniques. Biomed Signal Process Control 72:103341
10. Anilkumar KK, Manoj VJ, Sagi TM (2022) Automated detection of b cell and t cell acute lymphoblastic leukaemia using deep learning. IRBM 43(5):405–413

11. Shafique S, Tehsin S (2018) Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks. *Technol Cancer Res Treat* 17:1533033818802789
12. Thanh TTP, Vununu C, Atoev S, Lee SH, Kwon KR (2018) Leukemia blood cell image classification using convolutional neural network. *Int J Comput Theory Eng* 10(2):54–58
13. Shahin AI, Guo Y, Amin KM, Sharawi AA (2019) White blood cells identification system based on convolutional deep neural learning networks. *Comput Methods Programs Biomed* 168:69–80
14. Das PK, Meher S (2021) Transfer learning-based automatic detection of acute lymphocytic leukemia. In: National conference on communications (NCC). IEEE, Kanpur, India, pp 1–6
15. Das PK, Meher S (2021) An efficient deep convolutional neural network based detection and classification of acute lymphoblastic leukemia. *Expert Syst Appl* 183:115311
16. Mohammed KK, Hassanien AE, Afify HM (2023) Refinement of ensemble strategy for acute lymphoblastic leukemia microscopic images using hybrid CNN-GRU-BiLSTM and MSVM classifier. *Neural Comput Appl* 35:17415–17427
17. Dhalla S, Mittal A, Gupta S (2024) LeukoCapsNet: a resource-efficient modified CapsNet model to identify leukemia from blood smear images. *Neural Comput Appl* 36:2507–2524
18. Labati RD, Piuri V, Scotti F (2011) All-IDB: the acute lymphoblastic leukemia image database for image processing. In: 18th IEEE international conference on image processing. IEEE, pp 2045–2048
19. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *ArXiv preprint arXiv:1409.1556*
20. Mirjalili S (2016) SCA: a sine cosine algorithm for solving optimization problems. *Knowl-Based Syst* 96:120–133
21. Jain A, Nandakumar K, Ross A (2005) Score normalization in multimodal biometric systems. *Pattern Recogn* 38(12):2270–2285
22. Chollet F (2015) Keras: deep learning library for theano and tensorflow. <https://keras.io/k>
23. Sutskever I, Martens J, Dahl G, Hinton G (2013) On the importance of initialization and momentum in deep learning. In: International conference on machine learning. PMLR, pp 1139–1147
24. Guo Y, Li JY, Zhan ZH (2020) Efficient hyperparameter optimization for convolution neural networks in deep learning: a distributed particle swarm optimization approach. *Cybern Syst* 52(1):36–57
25. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. *Adv Eng Softw* 69:46–61
26. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, pp 770–778
27. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, pp 4700–4708
28. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) Mobilenets: efficient convolutional neural networks for mobile vision applications. *ArXiv preprint arXiv:1704.04861*

# Computer Vision and Deep Learning-Based Model for Detecting Spoofed Faces in Images



Gayathri P Salian, Manasa K Rao, and M. Rashmi

**Abstract** In the modern era of smart applications, video data is critically important in various contexts. In most of these applications, cameras are frequently incorporated to facilitate authentication. As a result, face recognition is the biometric method most frequently employed to authenticate users in these applications. The vulnerability of face recognition systems to spoofing attacks grows in tandem with their increased usage. As a result, robust countermeasures are required. This paper presents an approach to face anti-spoofing through transfer learning and YOLOv8 optimization. Additionally, a custom dataset was constructed using images obtained from web cameras and an existing dataset to assess the proposed work's real-time effectiveness. The proposed approach also adds a blurriness threshold during image capture to improve performance. With a mean Average Precision (mAP50) of 0.975, the experimental outcomes highlight the model's effectiveness in detecting face spoofing.

**Keywords** Biometric · Deep learning · Face spoofing · Face recognition · Smart applications · Surveillance

## 1 Introduction

Recently, face recognition systems have become essential for user identification, security, and comfort in life. However, due to their widespread implementation, these systems have become susceptible to various security threats, including spoofing attacks. Face anti-spoofing is a critical defensive mechanism that distinguishes genuine facial attributes from deceptive imitations in images. In print attack [1, 2],

---

G. P. Salian (✉) · M. K. Rao · M. Rashmi

Department of Data Science and Computer Applications, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, Karnataka, India  
e-mail: [gayathripsalian2001@gmail.com](mailto:gayathripsalian2001@gmail.com)

M. K. Rao  
e-mail: [raomanasa133@gmail.com](mailto:raomanasa133@gmail.com)

M. Rashmi  
e-mail: [rashmi.m@manipal.edu](mailto:rashmi.m@manipal.edu)

an attacker can exploit a valid user's photo printed on paper or shown on a digital device to enter the system. Also, an attacker using video/replay attacks leverages the typical movements of a right user that are recorded on camera [3].

Sophisticated spoofing attacks have exposed the vulnerabilities of cutting-edge facial recognition systems used in various applications. Existing facial recognition systems are gravely compromised by several attacks, which include print, replay, makeup, and 3D-mask-based assaults [4].

Face anti-spoofing emerges as a critical component in response to these challenges, enhancing the dependability and security of facial recognition systems. Face anti-spoofing techniques improve the accuracy and adaptability of face verification procedures and also find utility in a wide range of sectors by emphasizing detecting and thwarting fraudulent exploits. The significance of facial recognition technology is further proven by its contribution to the assurance of public trust and adherence to security standards. Face anti-spoofing is a critical safeguard against unauthorized access and identity fraud in the ever-evolving realm of biometric security, owing to its imperceptible characteristics that improve user experiences [5].

Several methods were proposed in the literature for authentication purposes based on different types of input data [6, 7]. Generally, any technique capable of autonomously differentiating between authentic biometric characteristics inputted into the sensor and artifacts manufactured synthetically incorporating said characteristic is regarded as an anti-spoofing method. Face anti-spoofing methods range from machine learning to advanced deep learning-based techniques. Depth analysis [8] employs three-dimensional depth data to differentiate planar presentations.

Motion analysis [9] introduces dynamism to the authentication process by monitoring facial movements. Texture analysis [10] introduces complexity by examining the intricacies of the skin. Supervised learning, as demonstrated in [11], distinguishes between authentic and fabricated features by applying SVMs and CNNs to massive datasets. As an alternative to explicit labeling, unsupervised methods such as [12] investigate inherent differences to identify novel deception techniques.

The prevalence of computer usage has escalated across various domains, augmenting the vulnerability to attacks, most notably face spoofing. Consequently, enhancing the functionality of anti-spoofing systems remains an arduous area of research [13, 14]. The key contributions of this work are as follows:

- The meticulous adaptation and optimization of YOLOv8 [15] to specifically address the intricacies of face anti-spoofing.
- A model capable of operating in real-time that is specifically engineered to surpass potential deception barriers when authentic facial features are present.
- To enhance the capability of the model in confronting recognition systems to identify and counteract spoofing attempts, a variety of data augmentation techniques were utilized, including image inversion and illumination condition adjustments.
- Incorporating a dual dataset consisting of both self-captured images and a variety of external datasets significantly enhances the model's adaptability and resilience to real-world scenarios.

Following this is the structure of the subsequent sections of this paper: A summary of relevant research in face anti-spoofing is presented in Sect. 2. Section 3 presents the proposed methodology, detailing the modification of YOLOv8 to enable face anti-spoofing. Section 4 covers the experimental configuration, results, and comprehensive analyses. In summary, Sect. 5 provides concluding remarks encompassing significant discoveries, noteworthy contributions, and recommendations for further investigation concerning face anti-spoofing utilizing YOLOv8.

## 2 Related Works

Several works are proposed for face anti-spoofing using handcrafted [16] and deep learning based methods [17]. A face recognition method was proposed by Balamurali K et al. [3] that incorporates denoising and utilizes two color spaces (YCbCr and CIELUV) to improve the discriminatory capability of embeddings extracted from VGG-Face. Using a Support Vector Classifier (SVC), the strategy exhibits potential in controlled environments. However, a significant limitation surfaced in the context of real-time applications: inadequate illumination substantially affects classification accuracy. The method's susceptibility to fluctuations in illumination underscores the need for additional research and possible enhancements to guarantee its dependability in practical, real-life situations. Yousef et al. [17] employed a patch-based CNN to extract local features and a depth-based CNN to generate depth maps for identifying real and spoof face images. In this case, the patch-based CNN was specifically utilized to augment the training data and preserve the original image's native resolution. Further, the classification was performed using features extracted from the depth map through the depth-based CNN. Also, they performed fusion analysis of different data modalities.

To simulate digital medium-based face spoofing attacks, Xiao Yang et al. [18] proposed a data collection solution and synthesis technique that allowed the acquisition of vast amounts of reflective, real-world training data. Leveraging the Spatio-Temporal Anti-Spoof Network (STASN), their approach significantly outperforms state-of-the-art methods on public face anti-spoofing datasets. However, the potential practical significance and impact of models trained on datasets that do not accurately depict real-world scenarios is a disadvantage.

Rathgeb et al. [19] suggested that M-PAD systems employ classifiers based on machine learning trained on synthetic images. Although developing an end-to-end M-PAD system using deep learning is possible, it requires a significant amount of training data. The limited availability of high-quality bona fide face images in databases poses a challenge for preventing overfitting. The constraints of the datasets might compromise the effectiveness of the proposed systems despite the implementation of diverse assault detection schemes and feature extractors. Addressing these challenges through access to larger-scale face databases is essential for achieving robust makeup presentation attack detection.

The video summarization method for face anti-spoofing tasks, proposed by Usman Muhammad et al. [20], is motivated by visual saliency theory. The authors intend to improve the performance of deep learning models through visual saliency. By extracting saliency information from the differences between the outputs of the Laplacian and Wiener filters on the source images, it is possible to determine which regions within each frame are the most visually prominent. Despite this, the approach may fail to capture every minute detail, particularly in the case of lengthier videos, as acknowledged in the paper. Jukka et al. [21] presented a method for processing input images that utilizes face and upper body detectors. Once the upper body has been identified, the image is subjected to spoofing medium detection to distinguish between genuine and spoof faces ultimately.

Previous investigations have employed various models, each with distinct merits and drawbacks. These models include denoising methods, patch-based and depth-based CNNs, synthetic image generation, and visual saliency techniques. This paper intends to introduce a novel methodology for face anti-spoofing by incorporating YOLOv8. In contrast to the anchor-dependent models utilized in previous research, YOLOv8 functions as an anchor-free model, making direct predictions regarding the center of an object. This distinctive characteristic offers a pathway to address present difficulties and enhance the robustness of face anti-spoofing techniques in practical situations.

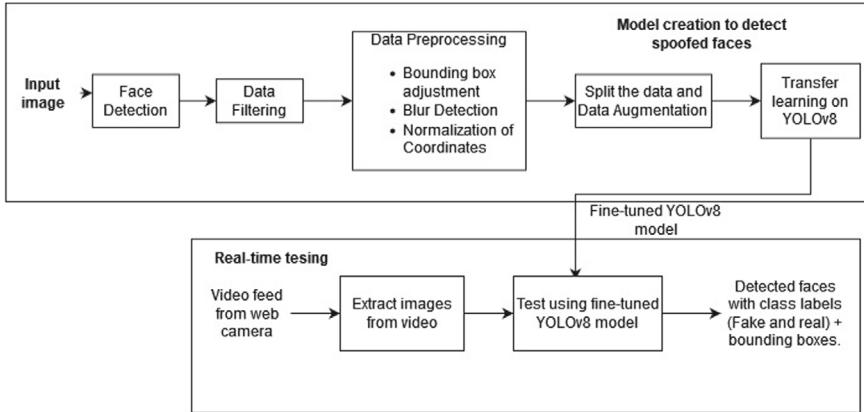
### 3 Proposed Methodology

The workflow illustrating the proposed method for face spoofing detection is depicted in Fig. 1. It focuses primarily on face spoof detection and data collection. The methodology employed to identify face spoofs is additionally subdivided into a number of sub-steps. Also, a series of steps are applied to collect the required data from real-time videos. A detailed explanation is provided below.

#### 3.1 Data Collection

The proposed method incorporates both custom-collected and data from existing datasets. A computer system with a webcam was utilized to gather the customized data. The OpenCV and CVZone libraries are employed to perform video processing and face detection.

**Face Detection** The dimension of real-time video frames captured using a webcam is set to (640,480). To achieve accurate face detection in every video frame, we utilized the FaceDetector module provided by CVZone. The bounding box coordinates of detected faces are extracted and saved to a text file.



**Fig. 1** Proposed model for face anti-spoofing.

**Data Filtering and Labeling** A data filtration scheme was implemented to eliminate images of poor quality. Faces with confidence scores surpassing a predetermined threshold are retained via confidence-based filtering. Bounding box adjustments and blur detection mechanisms optimize face coverage and clarity. The dataset is meticulously labeled, distinguishing between real and fake faces using normalized coordinates in YOLOv8 format.

**Data Preprocessing** In data preprocessing, bounding box adjustments optimize face coverage while preventing negative values. Blur detection is applied to a region of interest along with face detection. Additionally, coordinates are normalized to ensure they do not exceed 1.

**Data Augmentation** The number of samples in the training part of the dataset is increased by adding images of varying quality using various data augmentation techniques. The image library is utilized to accomplish the augmentation by implementing linear contrast adjustments and horizontal rotations with a 50% probability. By applying the augmentation pipeline to each image in a sequential fashion, the diversity of the dataset is increased. Augmented images passing confidence and blur checks are promptly saved with timestamped names, contributing to a comprehensive and robust dataset for effective model training and evaluation. Additionally, the dataset contains the bounding box coordinates of the identified faces in augmented images. Crucially, during the augmentation process, meticulous labeling is conducted for the augmented images. These labels accurately capture the transformations applied to the original images, ensuring alignment between visual content and annotations. Furthermore, file names are timestamped to guarantee uniqueness, optimizing the organization and traceability of the augmented dataset. Algorithm 1 summarizes the various steps involved in data collection.

**Algorithm 1:** Algorithm for selecting images from a spontaneous video stream

---

**Input:** Video Stream from web camera  
**Output:** Non-blur image frames

- 1 Initialize: Set parameters for FaceDetector with CVZone.  
 Extract first image frame from video stream  
 Pre-process the image frame  
 $image\_frame = \text{first\_frame\_of\_video}$   
 $c\_threshold = 0.8$   
 $b\_threshold = 35$   
**while**  $image\_frame$  **do**
- 2     Preprocess  $image\_frame$ .  
 Detect faces using FaceDetector.  
 Get bounding boxes for each face.  
 $Confidence = \text{Check\_Confidence}.$   
**if**  $confidence \geq c\_threshold$  **then**  
- 3         Call blur detection function.  
 $blur\_value = \text{Apply\_blur\_detection}(image\_frame)$   
**if**  $blur\_value \geq b\_threshold$  **then**  
- 4             Add  $image\_frame$  to "Non-Blur Images".  
 Normalize bounding box
- 5         **else**  
- 6             Discard  $image\_frame$
- 7         **end**
- 8     **else**  
- 9          $image\_frame = \text{read next\_frame\_from\_video}.$
- 10    **end**
- 11 **end**

---

### 3.2 Face Spoof Detection

The proposed method comprises a number of sub-modules. It begins with data processing and continues with YOLOv8 classification fine-tuning and real-time video testing. The detailed explanation is given below. **Data Preprocessing** The proposed method used the FaceDetectionModule from CVZone for face detection in each image. The confidence-based filtering is used to selectively retain the faces surpassing confidence thresholds. The current dataset photos are preprocessed using the same methods used for bespoke dataset preparation. The different steps are as follows.

- Bounding Box Adjustment Adjust detected bounding boxes to ensure comprehensive face coverage. Apply offsets to avoid negative values.
- Blur Detection: Extract a region of interest (ROI) representing the detected face. Use the Laplacian operator to calculate the blur value. Consider faces with blur values below a specified threshold as non-blurry.
- Normalization of Coordinates: Normalize coordinates to values between 0 and 1.

**Fine-Tuning YOLOv8** Deep learning algorithms, such as the YOLO series [22], have successfully balanced speed and accuracy in object detection tasks [23]. In the

proposed method, YOLOv8, a deep learning model known for its sophisticated object detection capabilities that are ideal for real-time applications, is utilized. YOLOv8 employs the capabilities of deep CNNs, specifically Darknet [24], to extract features. The model, trained on large corpora such as ImageNet [25, 26], has been fine-tuned for the proposed task of recognizing spoofed faces. Pre-trained weights are utilized to initialize the model during training to ensure that learned features are retained and that the model can adapt to dataset-specific attributes. Subsequently, the refined YOLOv8 model is used in real-time evaluation to distinguish between authentic and spoofed faces in images obtained from videos captured with a webcam..

The training is configured for 50 epochs, with a batch size of 16, input image size of  $640 \times 640$ , and the AdamW optimizer [27] with a learning rate of 0.001667 and momentum of 0.9. Also, applied early stopping after 25 epochs. To expedite the procedure, the proposed method used Automatic Mixed Precision (AMP), which is memory-efficient. It also enhances efficiency by adjusting precision in response to computational demands.

**Real-Time Testing** The fine-tuned YOLOv8 model is used to test the real images. The images are extracted from video captured using a web camera. The extracted frames are preprocessed and fed to the fine-tuned YOLOv8 to find spoofed and real images.

## 4 Experiments, Results and Analysis

### 4.1 Dataset

The proposed approach is tested on two datasets, namely, “Large Crowd Collected Dataset” (LCCD) [28] and a custom dataset, which combines the LCCD dataset with images collected by our team. Following an initial filtration process based on a blur threshold, the resulting images were preserved for further analysis. The LCCD dataset, containing images captured in large crowd scenarios, comprised 2588 non-blurred images, while the custom dataset, incorporating additional images collected by our team, contained 3338 non-blurred images. A range of data augmentation methods have been applied to the training sets of both datasets, including median blur, grey-scale conversion, and Contrast Limited Adaptive Histogram Equalization (CLAHE). As a result, the LCCD dataset is expanded to 3622 images, while the custom dataset is increased to 3750 images. Following this, a blurriness-based filtration was carried out to ensure that the dataset contained only high-quality images. Both datasets were divided into three subsets-train, test, and validation-in a ratio of 7:2:1. The validation sets for the LCCD and custom datasets comprised 518 and 666 images, respectively. Meanwhile, the LCCD and custom dataset test sets comprise 259 and 334 samples.

## 4.2 Results and Analysis

The proposed method is evaluated using Mean Average Precision (mAP), Precision, Recall, and fitness evaluation metrics. mAP50 is a metric used to measure the average precision of an object detection across different classes when considering only predictions with Intersection over Union (IoU) overlap of at least 50% with the ground truth bounding boxes. mAP50-95 (mean Average Precision from 50 to 95) is another metric extending the evaluation range beyond just 50% IoU to include IoU values from 50% to 95%. mAP50-95 provides a much better assessment of the model's performance across a wider range of IoU thresholds. This metric provides a whole picture of the model's accuracy at various levels of overlap between predicted and ground truth bounding boxes.

Table 1 shows the performance of the proposed model concerning various evaluation metrics on the LCCD and Custom dataset. This indicates that on the LCCD dataset, the performance of the proposed approach improved after the data augmentation. The proposed model showed improved mAP50 performance on the custom dataset but decreased slightly on mAP50-95. However, the proposed model showed promising performance on custom datasets, which include images captured through web cameras.

The sample outcomes of real-time assessment are illustrated in Figs. 2a and b. Figure 2a depicts a real person in front of a web camera while holding a photograph of a different individual. The proposed method distinguishes precisely between authentic and fake faces. An image of a person is presented to the camera in Fig. 2b; the proposed model identified it as a fake face. Figures 3, 4, and 5 depicts the change in Precision, Recall, and mAP50-95 in various steps on both datasets.

The experiments utilized Ultralytics YOLOv8 version 1.9 implemented in Python 3.10.13 with PyTorch version 2.1.2. A Tesla T4 GPU with 15102MiB memory was employed for CUDA-based acceleration. The Tesla T4 GPU, on Kaggle offers a stable infrastructure for machine learning tasks to facilitate rapid model iteration and optimization, thereby ensuring the necessary high computational performance. The summary of specification of various parameters are reported in Table 2.

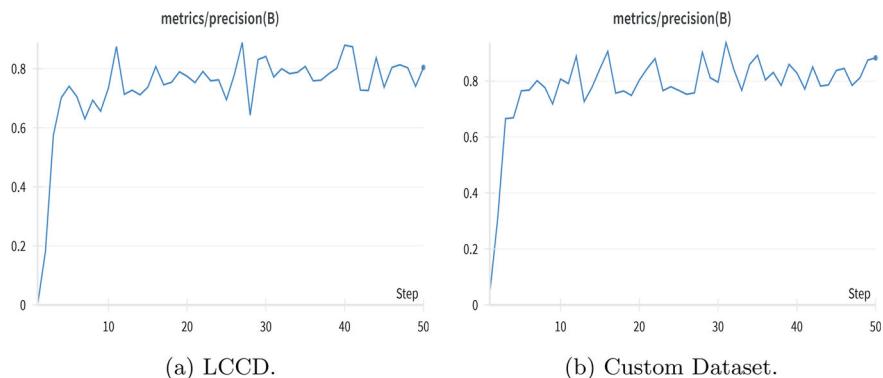
**Table 1** Performance metrics of different datasets with and without augmentation

Dataset	Augmentation	Precision (%)	Recall (%)	mAP50 (%)	mAP50-9 (%)	Fitness (%)
LCCD	No	80.45	89.36	97.42	93.07	93.50
LCCD	Yes	87.52	90.62	98.13	94.27	94.66
Custom	No	88.31	90.68	96.66	91.58	92.08
Custom	Yes	88.59	94.27	96.83	87.64	88.56



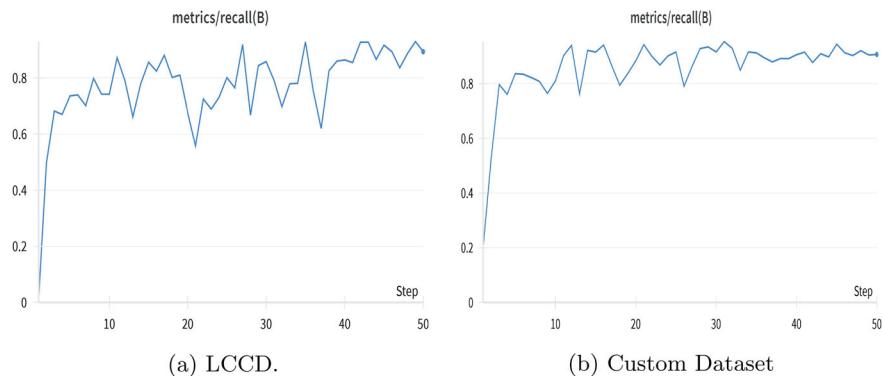
(a) Multiple face spoof detection.

(b) Single face spoof detection

**Fig. 2** Sample test images and detection results

(a) LCCD.

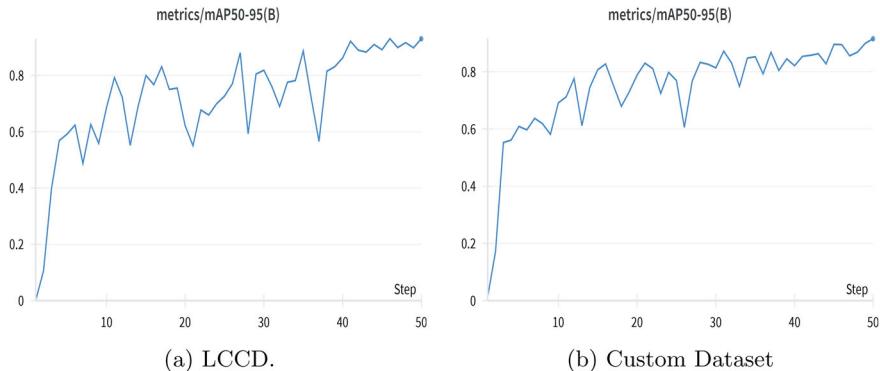
(b) Custom Dataset.

**Fig. 3** Change in precision versus steps

(a) LCCD.

(b) Custom Dataset

**Fig. 4** Change in recall versus steps



**Fig. 5** Change mAP50-95 versus steps

**Table 2** Specification of parameters and experimental environment used in the proposed method.

Parameter name	Specification
Software framework and environment	Ultralytics YOLOv8 version 1.9 implemented in Python 3.10.13 with PyTorch version 2.1.2
Computation system	Tesla T4 GPU offered by Kaggle
No. of epochs	50
Batch size	16
Input image size	640 × 640
Optimizer	AdamW
Learning rate	0.001667
Momentum	0.9

## 5 Conclusion

This paper proposes an enhanced face anti-spoofing strategy that optimizes YOLOv8 for greater precision and speed. The refined methodology strategically leverages real-time object detection to discern authentic facial traits from potential spoofing attempts. The working of the proposed methodology in real-time images is also experimented with by creating a custom dataset. Experimental outcomes, notably a mAP50 of 0.9753, underscore the model's efficacy against face spoofing. This research contributes substantially to enhancing anti-spoofing measures, suggesting future directions for bolstering security in facial recognition systems.

The future work involves developing more improved deep learning-based models that adapt to real-world scenarios in detecting spoofed faces and improving the performance of face recognition systems. To improve facial recognition system security, create more realistic datasets that correctly portray real-world settings in different scenarios.

## References

1. Yu Z, Qin Y, Li X, Zhao C, Lei Z, Zhao G (2023) Deep learning for face anti-spoofing: A survey. *IEEE Transactions On Pattern Analysis And Machine Intelligence*. 45:5609–5631. <https://doi.org/10.1109/TPAMI.2022.3215850>
2. Liu A, Tan Z, Yu Z, Zhao C, Wan J, Lei Y, Zhang D, Li S, Guo G (2023) Fm-vit: Flexible modal vision transformers for face anti-spoofing. *IEEE Trans Inf Forens Secur*. <https://doi.org/10.1109/TIFS.2023.3296330>
3. Balamurali K, Chandru S, Razvi M (1917) Kumar V (2021) Face spoof detection using vgg-face architecture. *J Phys: Conf Ser* 1:012010
4. Yu Z, Cai R, Li Z, Yang W, Shi J, Kot A (2024) Benchmarking joint face spoofing and forgery detection with visual and physiological cues. *IEEE Trans Dependable Secure Comput* 1–15. <https://doi.org/10.1109/TDSC.2024.3352049>
5. Li H, Li W, Cao H, Wang S, Huang F, Kot AC (2018) Unsupervised Domain Adaptation for Face Anti-Spoofing. *IEEE Trans Inf Forens Secur* 13(7):1794–1809. <https://doi.org/10.1109/TIFS.2018.2801312>
6. Debnath S, Roy P (2020) User authentication system based on speech and cascade hybrid facial feature. *Int J Image Graph* 20(03):2050022
7. Debnath S, Ramalakshmi K, Senbagavalli M (2022) Multimodal authentication system based on audio-visual data: a review. In: 2022 international conference for advancement in technology (ICONAT), Goa, India, pp 1–5. <https://doi.org/10.1109/ICONAT53423.2022.9725889>
8. Wu Z, Cheng Y, Yang J, Ji X, Xu W (2023) DepthFake: spoofing 3D face authentication with a 2D photo. In: 2023 IEEE symposium on security and privacy (SP), pp 917–91373
9. Qi H, Wu C, Shi Y, Qi X, Duan K, Wang X (2023) A real-time face detection method based on blink detection. *IEEE Access* 11:28180–28189. <https://doi.org/10.1109/ACCESS.2023.3257986>
10. Wang Z, Yu Z, Wang X, Qin Y, Li J, Zhao C, Liu X, Lei Z (2023) Consistency regularization for deep face anti-spoofing. *IEEE Trans Inf Forens Secur* 18:1127–1140
11. Guo S, Chen S, Li Y (2016) Face recognition based on convolutional neural network and support vector machine. In: 2016 IEEE international conference on information and automation (ICIA), pp 1787–1792
12. Khairnar S, Gite S, Kotecha K, Thepade S (2023) Face liveness detection using artificial intelligence techniques: a systematic literature review and future directions. *Big Data Cognit Comput* 7:37. <https://doi.org/10.3390/bdcc7010037>
13. Malviya S, Kumar P, Namasudra S, Tiwary US (2022) Experience replay-based deep reinforcement learning for dialogue management optimisation. *Trans Asian Low-Resour Lang Inf Process*. <https://doi.org/10.1145/3539223>
14. Das S, Namasudra S (2023) MACPABE: multi-authority-based CP-ABE with efficient attribute revocation for IoT-enabled healthcare infrastructure. *Int J Netw Manag* 33(3):e2200. <https://doi.org/10.1002/nem.2200>
15. Ultralytics YOLOv8 (2023). <https://github.com/ultralytics/ultralytics>
16. Das P, Hu B, Liu C, Cui K, Ranjan P, Xiong G (2019) A new approach for face anti-spoofing using handcrafted and deep network features. In: 2019 IEEE international conference on service operations and logistics, and informatics (SOLI), pp 33–38
17. Atoum Y, Liu Y, Jourabloo A, Liu X (2017) Face anti-spoofing using patch and depth-based CNNs. In: 2017 IEEE international joint conference on biometrics (IJCB), pp 319–328
18. Yang X, Luo W, Bao L, Gao Y, Gong D, Zheng S, Li Z, Liu W (2019) Face anti-spoofing: model matters, so does data. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 3502–3511. <https://doi.org/10.1109/CVPR.2019.00362>
19. Rathgeb C, Drozdowski P, Busch C (2020) Makeup presentation attacks: review and detection performance benchmark. *IEEE Access* 8:224958–224973. <https://doi.org/10.1109/ACCESS.2020.3044723>
20. Muhammad U, Oussalah M, Hoque M, Laaksonen J (2023) Saliency-based video summarization for face anti-spoofing. *ArXiv Preprint ArXiv:2308.12364*

21. Määttä J, Hadid A, Pietikäinen M (2011) Face spoofing detection from single images using micro-texture analysis. In: 2011 international joint conference on biometrics (IJCB), pp. 1–7
22. 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), pp 779–788
23. Lin B (2024) Safety helmet detection based on improved YOLOv8. IEEE Access 12:28260–28272. <https://doi.org/10.1109/ACCESS.2024.3368161>
24. Kumar Gaurav A, DarkNet. <https://www.opensourceforu.com/2021/04/darknet-the-open-source-framework-for-deep-neural-networks>
25. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) ImageNet: a large-scale hierarchical image database. In: IEEE conference on computer vision and pattern recognition, Miami, FL, USA, pp 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
26. Ultralytics ImageNet. <https://docs.ultralytics.com/datasets/classify/imagenet/>
27. Loschilov I, Hutter F (2017) Fixing weight decay regularization in adam. CoRR. <http://arxiv.org/abs/1711.05101>
28. Timoshenko D, Simonchik K, Shutov V, Zhelezneva P, Grishkin V (2019) Large crowdcollected facial anti-spoofing dataset. In: 2019 computer science and information technologies (CSIT), pp 123–126

# Explainable AI Insights into a Time Series Weather Prediction Model Using Stacked LSTM



T. H. Sunu Fathima and Binsu C. Kovoor

**Abstract** The need for accurate weather prediction models is on the rise, considering their potential to make impactful decisions in various sectors of society. Numerical Weather Prediction (NWP) models have excelled with accurate predictions when data was limited. In the Big Data era, the availability of a large amount of diverse weather data has been a challenge to the forecasting power of the traditional models. Deep learning entered the picture as an alternative that can learn accurate patterns from complex datasets. In this paper, we implement a model that predicts short-term temperature based on a Stacked Long Short-Term Memory (LSTM) on historical Radar data collected from Cochin University of Science & Technology. An Explainable AI using SHapley Additive exPlanations (SHAP) is implemented on the model to determine the effect of different features on the predicted values. The proposed model is compared with a single layer vanilla LSTM and the performance is evaluated based on Root Mean Square Error (RMSE), Mean Absolute Error (MAE), R<sup>2</sup> score and Mean Absolute Percentage Error (MAPE). The results showed that the model outperforms the vanilla LSTM and performs on par with state-of-the-art methods. The SHAP results identify features that have an upper hand in the final predictions.

**Keywords** Temperature prediction · Long short-term memory · XAI · SHAP

---

T. H. Sunu Fathima · B. C. Kovoor

Division of Information Technology, Cochin University of Science and Technology, Kochi, Kerala, India

e-mail: [sunufathima93@gmail.com](mailto:sunufathima93@gmail.com)

B. C. Kovoor

e-mail: [binsu@cusat.ac.in](mailto:binsu@cusat.ac.in)

T. H. Sunu Fathima

Department of Computer Science, Rajagiri College of Social Sciences (Autonomous), Kalamassery, Kerala, India

## 1 Introduction

Weather prediction is one of the major aspects of modern society that requires significant attention because of its ability to provide impactful insights into the varying behavior of the atmosphere. It is the science of interpreting patterns from past weather data, deciphering the relationship between different features and foreseeing future weather conditions. Over the years, many studies have been conducted related to the prediction of weather parameters such as temperature, pressure and precipitation.

Traditional weather prediction models are based on many mathematical equations and algorithms. Hence, they are referenced as NWP models [1]. The introduction of computers to the forecasting problem revolutionized and increased the reliability of these traditional systems. The collection of high-precision real-time data and the requirement of large computational power for processing this data make it difficult for these models to sustain themselves. Thus, machine learning algorithms evolved. Dynamical weather states were examined properly through a set of algorithms and the right patterns were extracted in order to make weather-related decisions [2]. Several machine learning algorithms have found application in various fields over the years. After a while, Time Series Prediction which forecasts future values based on the trend of historical data became popular. The non-linear nature of most of the weather datasets introduced the need for complex predictive architectures such as Artificial Neural Networks (ANN). The availability of diverse datasets at large volumes from various sources in the era of big data [3, 4] ignited the need for advanced architectures for prediction.

Deep Neural Network (DNNs) addressed the challenges of the availability of large datasets, the need for high computational power and the significance of learning features automatically from the datasets [5]. Convolutional Neural Networks (CNNs) are one of the most impressive artificial neural networks that have been discovered so far. Using many convolution layers to extract the features along with pooling layers and fully connected networks provides more accurate classification outcomes. Different ensemble models also provide accurate results in various applications of prediction and classification [6]. Recurrent Neural Networks (RNNs) [7] are one of the best-known networks for handling time series data. The temporal dependencies from previous sequences are captured and stored, which may influence the values at a future point in time. LSTM [8] is the most popular RNN that can capture the long-term dependencies from the dataset and eliminate the vanishing gradient problem.

Machine learning models act as a black box where the user is unable to explain the role of different features that led them to the results obtained. Explainable AI (XAI) plays a significant role in providing appropriate interpretations for different researchers working on their model. Given a set of input features and the predictions generated from them, XAI provides a set of procedures that explain the contribution of each of the input features and their impact on the outcome [9].

Our proposed work is based on a Stacked Long Short-Term Memory Network that has two hidden LSTM layers of 32 units each. It is implemented on the weather data from the Advanced Centre for Atmospheric Radar Research (ACARR), Cochin

University of Science & Technology. Atmospheric Weather Station installed on various locations on the campus is used to collect weather data. The proposed forecasting model is trained, considering the dependency between various weather parameters available from the radar dataset. The model is further used to forecast temperature values for the future. The characteristics of the dataset are studied by conducting different experiments on the weather data by varying the hyperparameters, and the model that provides the best results is selected for further tests. An LSTM with 32 hidden units trained at a learning rate of 0.002 in 15 epochs exhibited the best performance. The performance is evaluated based on four metrics: RMSE, MAE, R<sup>2</sup> score and MAPE. An XAI using SHAP is implemented on top of the model to interpret the influence of different weather parameters on the result. The results showed that parameters such as solar irradiance, pressure and temperature highly influenced the forecasting results.

The contributions of the proposed research work include.

- A short-term temperature prediction model on the weather data from Radar at Cochin University of Science & Technology based on a stacked LSTM network.
- The model is compared with a single layer LSTM network, and its performance is evaluated using metrics such as RMSE, MAE, R<sup>2</sup> score and MAPE.
- XAI using SHAP is implemented on the model that illustrates the impact of various input features on the final prediction.

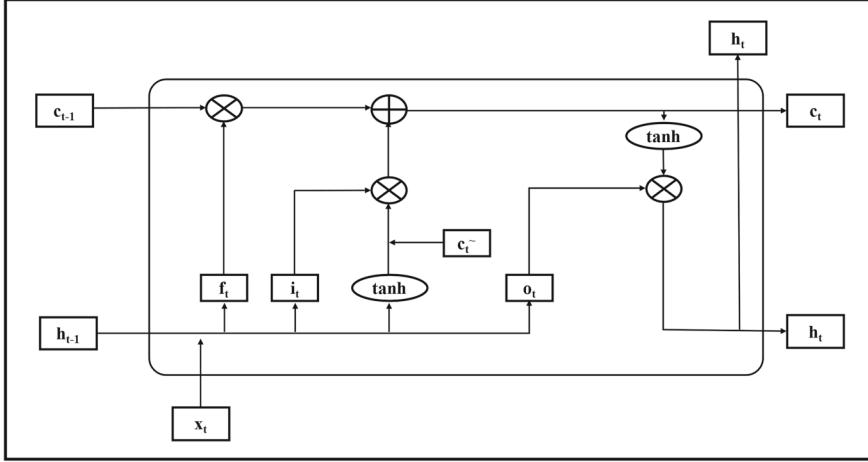
The rest of this paper is organized as follows: Sect. 2 summarizes the methods used for implementing the model. The proposed method is briefly introduced in Sect. 3. The performance metrics used to evaluate the model and the results obtained are discussed in Sect. 4. Section 5 concludes the research work and provides future directions.

## 2 Methods Used

### 2.1 Long Short-Term Memory

LSTM [8] models are RNNs [7] that are developed to address the vanishing gradient [10] problem and also introduce the concept of long-term dependencies [11] into the sequencing problem. It is based on the concept that the prediction of the next sequence in a sequencing problem depends not only on its immediate predecessor but also on the earlier sequences in the chain. The key components of an LSTM model are the forget gate, input gate, output gate, hidden state and cell state (see Fig. 1). The working of these gates and the memory cell determines which information is to be discarded from the further flow and which should be kept [12]. The mathematical calculations for finding the outputs of various gates are given in Eqs. (1)–(6):

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



**Fig. 1** An LSTM memory unit

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (4)$$

$$h_t = o_t * \tanh(c_t) \quad (5)$$

$$c_t = f_t * c_{t-1}, i_t * \tilde{c}_t \quad (6)$$

where  $h_t$  is the hidden state at time  $t$ ,  $f_t$  is the output of forget gate at time  $t$ ,  $o_t$  is the output of output gate at time  $t$ ,  $c_t$  is the cell state at time  $t$ ,  $c_{t-1}$  is the cell state at time  $t - 1$ ,  $x_t$  is the current input at time  $t$ ,  $h_{t-1}$  is the previous hidden state at time  $t - 1$ ,  $\tilde{c}_t$  is the candidate cell state,  $i_t$  is the input gate,  $W_f$ ,  $W_i$ ,  $W_c$  and  $W_o$  are the weight matrices and  $b_f$ ,  $b_{ci}$ ,  $b_o$  and  $b$  are the bias vectors.

## 2.2 SHapley Additive ExPlanations

The significance of providing valuable explanations for our predictions is increasing, considering the need for transparency in our model. XAI [9] is thus introduced to the world of deep learning. Many techniques are used to provide explanations. SHAP [13] is one of the most popular and powerful explanation technique. SHAP relies heavily

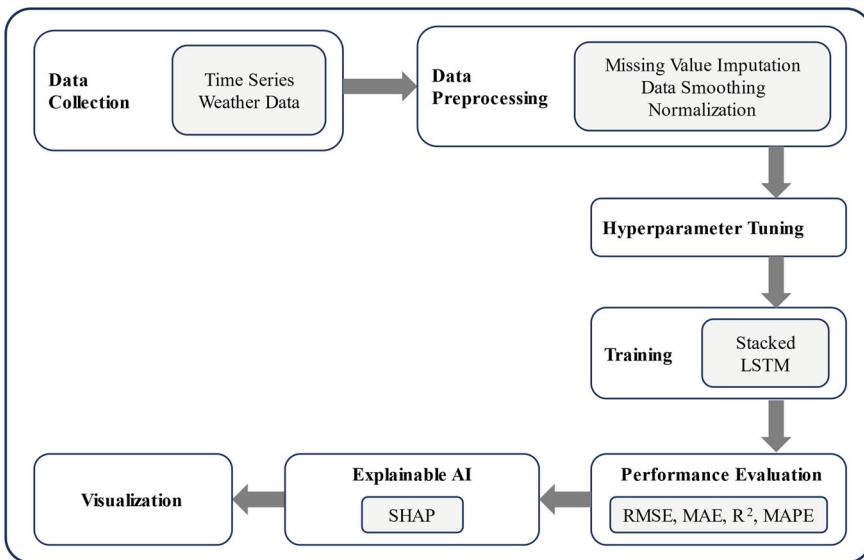
on the concept of the Shapley value, which is the average marginal contribution of a feature to the predicted value considering all possible combinations of the input features. For each prediction, it calculates the Shapley value, which is referred to as the local explanation. It then aggregates all the local explanations to create the global explanation [14]. This identifies the features with the most impact on the forecasts. SHAP values are calculated based on game theory [15] where the contribution of each player toward the final outcome is measured by Eq. (7):

$$g(z') = \emptyset_o + \sum_{j=1}^M \emptyset_j z'_j \quad (7)$$

where  $g$  is the model to be explained,  $z'_j \in \{0, 1\}^M$  is the binary coalition vector that determines whether the feature is present or not,  $M$  is the maximum coalition size, and  $\emptyset_j \in \mathbb{R}$  is the feature attribution for a feature  $j$ .

### 3 Proposed Method

The workflow of the proposed research work is given (see Fig. 2).



**Fig. 2** Workflow of the stacked LSTM-based temperature prediction model with SHAP

### 3.1 Data Collection

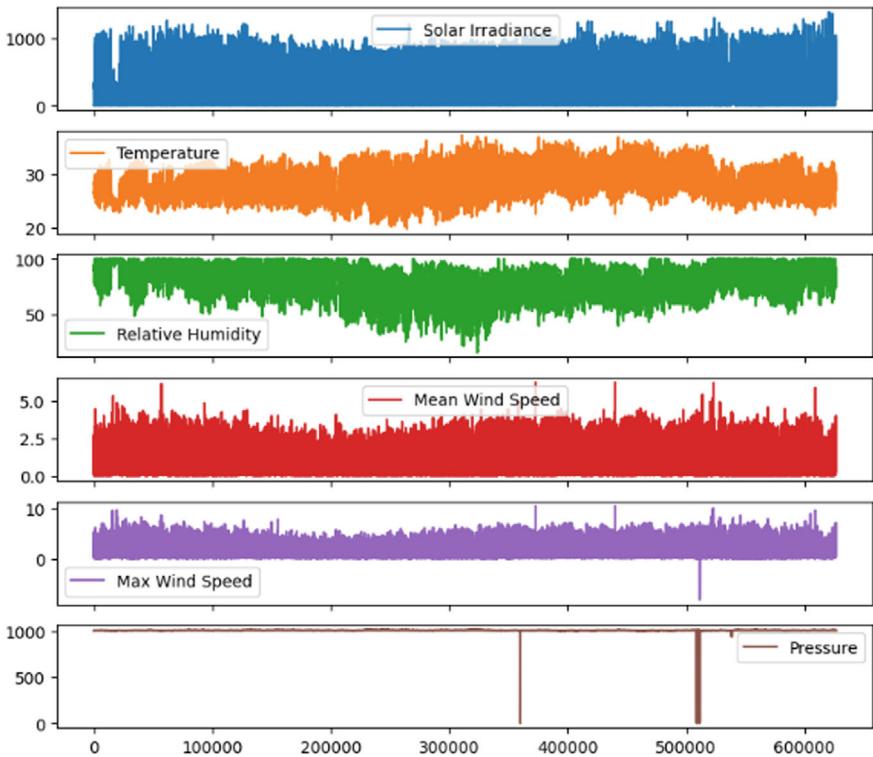
Weather information that is recorded and updated on a minute-by-minute basis from the Automatic Weather Station of ACARR at the Cochin University of Science & Technology campus is used for predicting short-term temperatures. The weather parameters are measured from the campus during the period from July 2022 to September 2023. It includes various weather data, out of which only Solar Irradiance, Temperature, Relative Humidity, Mean Wind Speed, Max Wind Speed and Pressure are chosen for our work. These six weather parameters are used to predict temperature values. The details of the input features are given in Table 1. A sample view of the dataset for the first five datapoints is given (see Fig. 3). The input dataset contains 6,26,595 datapoints. The data from the first five days is used to predict the temperature values for the next day. The input features are plotted to illustrate the range of values measured (see Fig. 4).

**Table 1** Description of input features selected from ACARR weather dataset

Sl. no.	Feature	Unit	Description
1	Solar irradiance	W/m <sup>2</sup> (Watts per square meter)	Average hourly solar irradiance (Solar radiation flux density)
2	Temperature	Degree Celsius	Average hourly air temperature
3	Relative humidity	Percentage	Relative humidity
4	Mean wind speed	m/s (Meter per second)	Mean wind speed
5	Max wind speed	m/s (Meter per second)	Maximum wind speed
6	Pressure	Millibar (mBar)	Barometric pressure

	Solar Irradiance	Temperature	Relative Humidity	Mean Wind Speed	Max Wind Speed	Pressure
0	254.6	28.33	78.87	2.513	3.38	1003.405
1	297.6	28.27	80.10	1.738	2.57	1003.430
2	301.3	28.28	79.20	1.352	2.12	1003.507
3	308.5	28.29	79.20	0.761	2.11	1003.499
4	206.8	28.36	80.30	1.408	2.38	1003.643

**Fig. 3** Sample view of the ACARR weather dataset



**Fig. 4** A plot of input weather features from ACARR weather dataset

### 3.2 Data Preprocessing

**Imputing Missing Values.** The quality of predictions is affected by too many missing values in the time series dataset. It will introduce biases and reduce efficiency. Hence, it is important to find an alternative for replacing the missing values in different fields with suitable data. The temporal dependence in the values between consecutive time steps of time series data can be utilized to find a solution. There are many methods available for this purpose. The proposed work uses Spline Interpolation [16] since our weather dataset from ACARR contains multiple parameters that exhibit both trends and seasonality. This technique approximates the missing data by employing a piecewise polynomial curve along the points to extract non-linear similarity between the data values.

**Data Smoothing.** The statistical techniques employed to eliminate noise and outliers from the time series data are known as data smoothing. They provide better understandability of the dataset, especially if there are trends. The most popular Moving Average [17] method is used for smoothing our weather dataset. The time series data

can be smoothed by calculating the average of all the data values in a sliding window with a fixed size and replacing the points with the new value. The window continues to stride over the data values and perform the same operation for the next sequence.

**Normalization.** The multivariate time series analysis introduces the need for rescaling the data into a common range since all the weather parameters are measured in different ranges and units. Z-score normalization [18] is used for the proposed model. Data after normalization for a particular field A can be calculated by Eq. (8):

$$\bar{x} = \frac{x - \sigma_A}{A^-} \quad (8)$$

where  $\bar{x}$  is the normalized value,  $x$  is the original value,  $\sigma_A$  is the mean of A, and  $A^-$  is the standard deviation of A.

### 3.3 Training

The dataset is divided into training set and testing set. The training set consists of 501,276 datapoints which is 80% of the overall dataset. The testing set consists of 125,319 datapoints. To optimize the performance of a single layer LSTM network, experiments are conducted by varying the number of hidden units, the number of epochs and the learning rate. The number of hidden units used in the experiment is 8, 16 and 32. Each of them is trained at a learning rate of 0.001 and 0.002. The number of epochs used in the experiment is 10, 15 and 20. The batch size is kept at 64, and the Adam optimizer is used with mean square error as the loss function throughout the training. The model is evaluated based on performance metrics such as RMSE, MAE,  $R^2$  score and MAPE. Error values are recorded after every experiment. Table 2 shows the comparison of error values obtained during training with different hyperparameters. The best values under each metric are highlighted.

### 3.4 Stacked LSTM with SHAP Explainability

An LSTM with 32 units at a learning rate of 0.002 when trained with 15 epochs exhibited the best performance. The performance of this model can be further enhanced by stacking another LSTM with 32 units and introducing appropriate Drop out and Dense layers. An XAI with SHAP is implemented on the model in order to find out which features from the input set influence the output predictions the most. This information can be used in future stages of the model refinement to select only relevant features. The architecture of the proposed model is given (see Fig. 5).

**Table 2** Comparison of error values obtained during hyperparameter tuning

No. of hidden units	Learning rate	No. of epochs	RMSE	MAE	R <sup>2</sup> score	MAPE
8	0.001	10	2.622	2.041	-0.063	7.47
		15	2.766	2.180	-0.183	8.08
		20	2.648	2.052	-0.084	7.58
	0.002	10	2.162	1.599	0.277	5.85
		15	1.871	1.541	0.689	5.71
		20	1.931	1.653	0.543	6.02
16	0.001	10	2.626	2.033	-0.066	7.4
		15	2.881	2.237	-0.283	8.28
		20	2.716	2.151	0.141	8
	0.002	10	2.302	1.919	0.524	5.9
		15	1.173	0.901	0.745	3.34
		20	1.786	1.672	0.632	6.2
32	0.001	10	2.070	1.382	0.337	5.07
		15	1.948	1.202	0.413	4.36
		20	1.811	1.098	0.493	4.01
	0.002	10	1.686	1.020	0.560	3.75
		15	<b>0.960</b>	<b>0.272</b>	<b>0.857</b>	<b>0.95</b>
		20	1.145	0.437	0.797	1.59

## 4 Performance Analysis

### 4.1 Performance Metrics

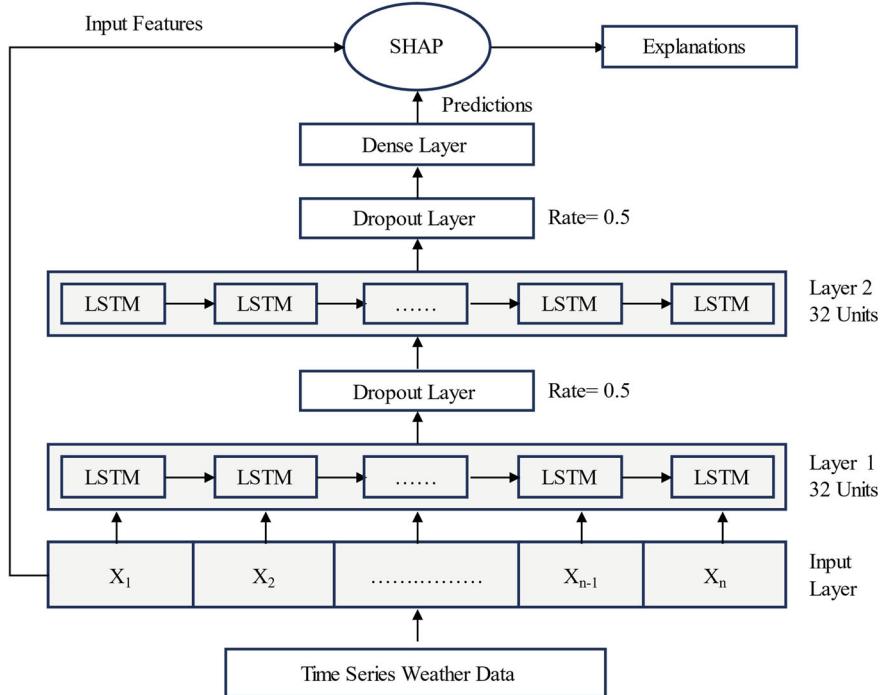
**Root Mean Square Error.** RMSE [19] can be calculated as the square root of the mean of the squared errors between predictions and actual values of a variable given by Eq. (9):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\text{Actual Value} - \text{Predicted Value})^2} \quad (9)$$

where n is the number of datapoints.

**Mean Absolute Error.** MAE [20] is calculated as the average variance between the predicted values and the actual values of a variable, as given by Eq. (10):

$$\text{MAE} = \frac{1}{n} \left( \sum_{i=1}^n |\text{Actual Value} - \text{Predicted Value}| \right) \quad (10)$$



**Fig. 5** Architecture of the stacked LSTM with SHAP explainability

where  $n$  is the number of datapoints.

**Coefficient of determination— $R^2$  score.**  $R^2$  score [21] is calculated as the measure of variation of an output dependent variable that can be predicted from an input independent variable, as given by Eq. (11):

$$R^2 = 1 - \frac{\sum_{i=1}^n (\text{Actual Value} - \text{Predicted Value})^2}{\sum_{i=1}^n (\text{Actual Value} - \text{Mean of Actual Value})^2} \quad (11)$$

where  $n$  is the number of datapoints.

**Mean Absolute Percentage Error.** MAPE [22] is a commonly used metric for regression problems that is defined as the average of absolute relative errors as given by Eq. (12):

$$\text{MAPE} = \frac{1}{n} \frac{|\text{Actual Value} - \text{Predicted Value}|}{\text{Actual Value}} \cdot 100\% \quad (12)$$

where  $n$  is the number of datapoints.

## 4.2 Results and Discussion

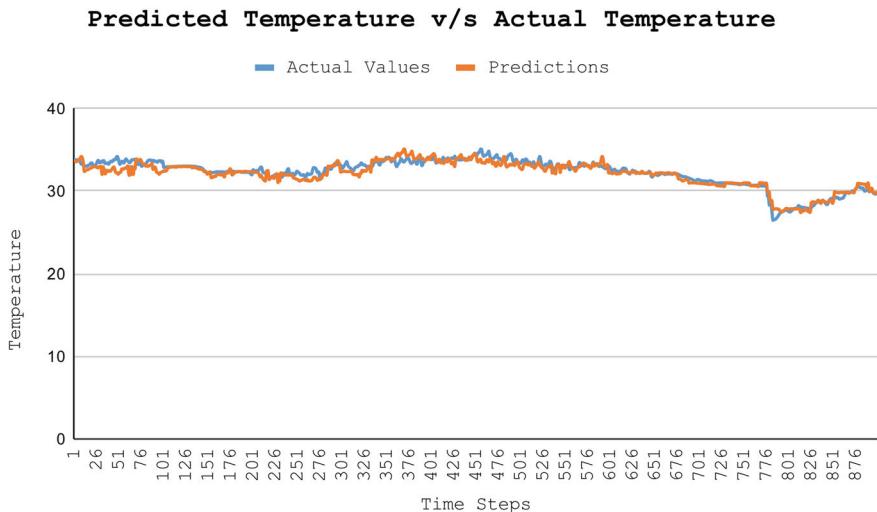
The proposed stacked LSTM model with SHAP explainability showed better performance compared to the single layer LSTM model. It produced a RMSE value of 0.762, a MAE of 0.127, an R<sup>2</sup> score value of 0.910 and a MAPE of 0.43. The comparison of the proposed model with a single layer vanilla LSTM is shown in Table 3.

The comparison of our model predictions with actual temperature values from the test set for a single day is shown in Fig. 6. The red plot indicates the temperature predictions generated by the stacked LSTM, and the blue plot indicates the original temperature at a particular time step that is given by the weather dataset. Data for an entire day is plotted on a minute-by-minute basis.

The results from the SHAP module highlighted the importance of each feature on the output values through different visualization techniques. Figure 7 shows the waterfall plot for one of the individual predictions. The value of the expected temperature, E[f(x)], for the corresponding datapoint is given at the bottom. x is the chosen datapoint, and f(x) is the predicted temperature. The length of the bar corresponding to each feature indicates the SHAP value of that particular feature in this prediction.

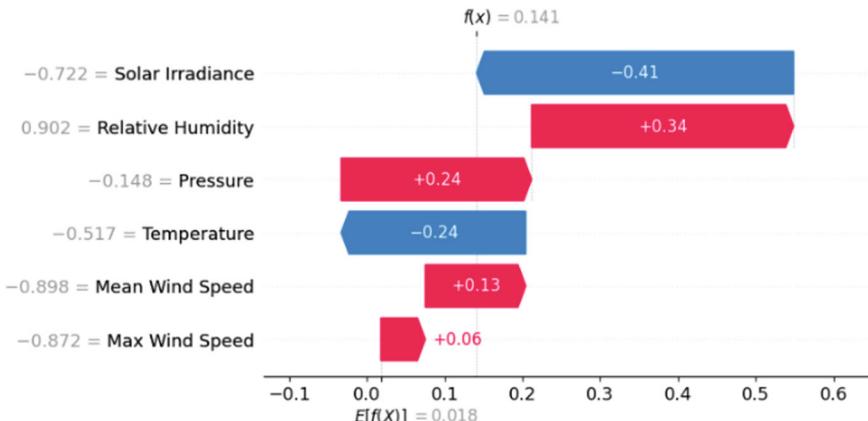
**Table 3** Comparison of the proposed Stacked LSTM with vanilla LSTM

Model	RMSE	MAE	R <sup>2</sup> score	MAPE
LSTM (32)	0.960	0.272	0.857	0.95
Stacked LSTM	0.762	0.127	0.910	0.43

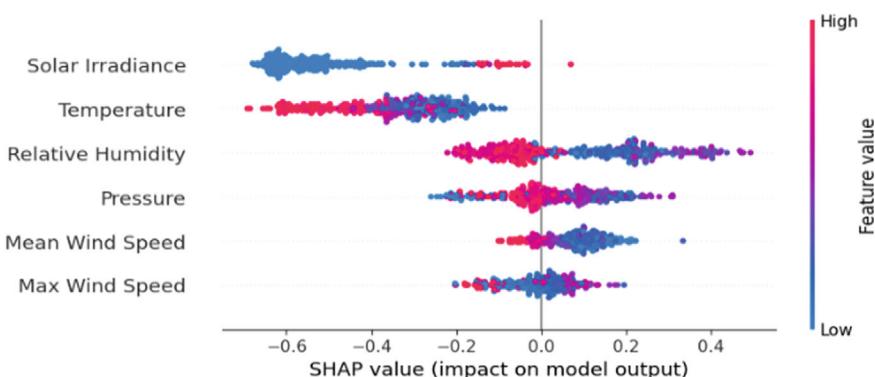


**Fig. 6** Predictions versus actual temperature for the first day in the test set

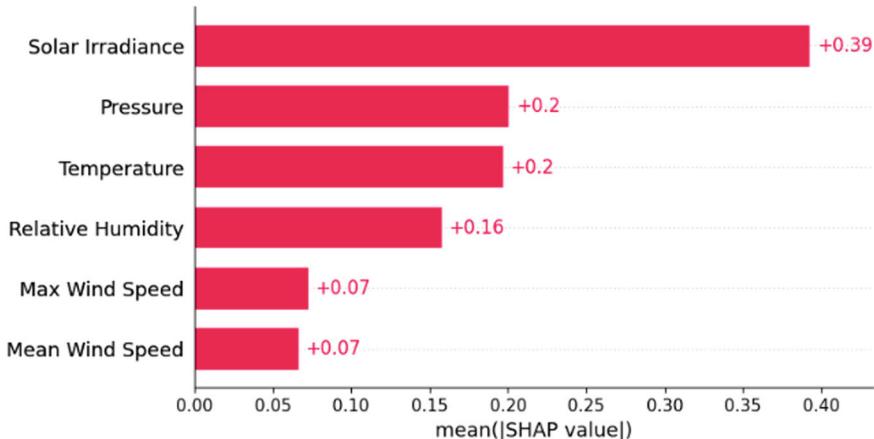
The contribution of each feature to the prediction is given by its absolute SHAP value. For the given datapoint, Solar Irradiance contributes the most to the outcome and Max Wind Speed has the least contribution. Figure 8 shows the swarm plot that illustrates different features of the dataset and their contribution to the overall predictions. In the swarm plot, red indicates high values and blue indicates low values. Based on the global contribution, features are listed. Low values of Solar Irradiance contribute negatively to the prediction, while low values of relative humidity contribute highly positively to the outcome. Figure 9 shows the bar plot, where each bar indicates the importance of that particular feature for the overall prediction. Features are listed according to their importance. Solar Irradiance has the highest impact, while mean wind speed has the lowest influence on the predictions.



**Fig. 7** Waterfall plot of SHAP values for a single datapoint from the weather dataset



**Fig. 8** Swarm plot for the weather dataset



**Fig. 9** Bar plot that illustrates the importance of each feature on the predictions

## 5 Conclusion

The proposed Stacked LSTM model with SHAP Explainability predicts short-term temperature values for time series weather data and explains the importance of each input feature on the output temperature value. The dependencies among various time steps from the weather information of the first five days can be modeled to predict the temperature for the next day. The training parameters of a vanilla LSTM model are optimized by conducting various experiments, and the performance of the model is enhanced by stacking another layer on the same model. This achieved a RMSE error of 0.762, a MAE value of 0.127, an R<sup>2</sup> score of 0.910 and a MAPE value of 0.43. The SHAP results proved that features such as Solar Irradiance, Pressure and Temperature largely affect the prediction. In the successive phases of model refinement, these explanations play an important role in selecting only the most relevant features and eliminating unnecessary features. It may further improve the performance and reduce computational overhead. In the future, the model can be extended to deploy hybrid architectures that may further increase the efficiency of forecasting and can be also used to predict long-term temperature values based on demand.

**Acknowledgements** This study was funded by RUSA 2.0 for Project ID: T4G.

## References

1. Bauer P (2015) The quiet revolution of numerical weather prediction. *Nature* 525(7567):47–55
2. Naveen L (2019) Atmospheric weather prediction using various machine learning techniques: a survey. In: 2019 3rd international conference on computing methodologies and communication (ICCMC). IEEE, Erode
3. Junqué de Fortuny E (2013) Predictive modeling with big data: Is bigger really better? *Big Data* 1(4):215–226
4. Namasudra S (2024) Enhanced neural network-based univariate time-series forecasting model for big data. *Big Data* 12(2):83–99
5. Schultz MG (2021) Can deep learning beat numerical weather prediction? *Philos Trans Royal Soc A: Math, Phys Eng Sci* 379(2194), Article 20200097
6. Agrawal D (2021) Ensemble algorithm using transfer learning for sheep breed classification. In: 2021 IEEE 15th international symposium on applied computational intelligence and informatics (SACI). IEEE, Timisoara, Romania
7. Tsantekidis A, Passalis N, Tefas A (2022) Recurrent neural networks. In: Deep learning for robot perception and cognition. Elsevier eBooks, pp 101–115
8. Hochreiter S (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
9. Xu F, Uszkoreit H, Du Y (2019) Explainable AI: a brief survey on history, research areas, approaches and challenges. In: Natural language processing and Chinese computing. Springer, pp 563–574
10. Hochreiter S (1998) The vanishing gradient problem during learning recurrent neural nets and problem solutions. *Internat J Uncertain Fuzziness Knowl-Based Syst* 06(02):107–116
11. Bengio Y (no date) The problem of learning long-term dependencies in recurrent networks. In: IEEE international conference on neural networks. IEEE, San Francisco, CA, USA
12. Hamad R. What is LSTM? Introduction to long short-term memory. Medium. <https://medium.com/@rebeeen.jaff/what-is-lstm-introduction-to-long-short-term-memory-66bd3855b9ce>. Accessed 8 Mar 2024
13. Molnar C (2022) Interpretable machine learning. <https://christophm.github.io/interpretable-ml-book/shap.html>
14. Lundberg S (2020) From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* 2(1):56–67
15. SAS Blogs. <https://blogs.sas.com/content/sasla/2021/04/26/kernel-shap-un-paso-adelante-serie-explicante/>. Accessed 17 Mar 2024
16. Wegman EJ (1983) Splines in statistics. *J Am Stat Assoc* 78(382):351–365
17. Raudys A (2013) Moving averages for financial data smoothing. *Commun Comput Inform Sci* 34–45
18. Adeyemo A (2019) Effects of normalization techniques on logistic regression in data science. *J Inform Syst Appl Res* 12(2):37–44
19. Hyndman RJ (2006) Another look at measures of forecast accuracy. *Int J Forecast* 22(4):679–688
20. Willmott C (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim Res* 30:79–82
21. Anderson-Sprecher R (1994) Model comparisons and  $r^2$ . *Am Stat* 48(2):113–117
22. De Myttenaere A (2016) Mean absolute percentage error for regression models. *Neurocomputing* 192:38–48

# Patient-Specific and Patient-Independent Seizure Prediction Using Ensemble Learning Technique



Ranjan Jana and Imon Mukherjee

**Abstract** Seizure poses life-threatening risks to individuals with epilepsy. Predicting seizures is challenging due to individual variations of EEG signal patterns. Channel selection is a crucial task in the effectiveness of seizure prediction devices. This work suggests utilizing a single EEG channel for designing small-sized, power-efficient, wearable prediction devices. The performances of all channels are not consistent when employing deep learning models. It leads to variations in the ranking of the channel's performance in each instance. In this work, each channel's performance is evaluated in five rounds, ultimately deriving a ranking of each channel based on the average performance across all rounds. The CNN-LSTM1D model is employed to calculate the performance of each channel. The channel securing the first rank is considered for efficient seizure prediction. Finally, an ensemble learning technique is employed that integrates CNN1D, DenseNet1D, and CNN-LSTM1D models to utilize a majority voting approach. Ensemble learning delivers superior performance compared to individual models using the CHB-MIT database. This research introduces effective strategies for both patient-specific and patient-independent seizure prediction. The patient-specific and patient-independent models provide a sensitivity of 0.9768 and 0.9425, and a specificity of 0.9681 and 0.9294, respectively. It outperforms existing state-of-the-art works.

**Keywords** Channel selection · CNN-LSTM · EEG signals · Epileptic seizure · Wearable device

---

R. Jana (✉) · I. Mukherjee  
Indian Institute of Information Technology, Kalyani, India  
e-mail: [ranjan\\_phd20@iiitkalyani.ac.in](mailto:ranjan_phd20@iiitkalyani.ac.in)

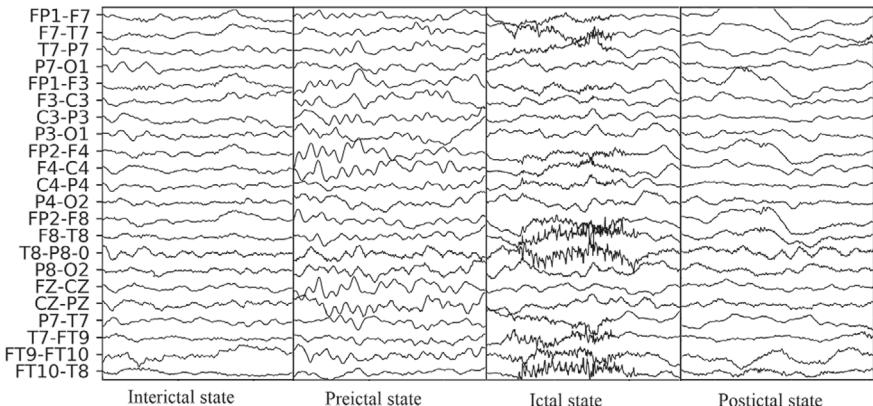
I. Mukherjee  
e-mail: [imon@iiitkalyani.ac.in](mailto:imon@iiitkalyani.ac.in)

R. Jana  
RCC Institute of Information Technology, Kolkata, India

## 1 Introduction

Abnormal electrical activity in the brain is indicated by an epileptic seizure, which leads to sudden loss of consciousness [1]. Individuals with epilepsy face life-threatening situations due to unpredictable seizure events. Seizures are characterized by symptoms like dizziness, jerky limb movements, and unconsciousness. Anti-seizure drugs can control seizures, though these drugs are effective for 70% of epilepsy patients [2]. Analyzing electroencephalogram (EEG) signals is a commonly used method for predicting seizures in advance [3]. The EEG data are collected from the scalp using metal electrodes. Epilepsy patients' states are categorized into four: ictal (during a seizure), pre-ictal (before seizure), post-ictal (after seizure), and inter-ictal (during normalcy) [2]. EEG signal patterns vary across these states, as illustrated in Fig. 1. Identifying pre-ictal states is essential for seizure prediction [4]. Therefore, inter-ictal and pre-ictal state classification is necessary for developing seizure prediction methods. Seizure prediction methods can be implemented in two ways: patient-specific and patient-independent [5]. EEG signal patterns are unique for each individual. Therefore, designing an effective patient-independent seizure prediction method is a challenging task [3]. Manual seizure prediction using EEG signals involves visual inspection and analysis of EEG patterns to anticipate the seizures in advance [3]. Capturing the continuous electrical activity in the patient's brain through ongoing EEG monitoring poses a challenging task for medical practitioners. Deep learning plays a significant role in seizure prediction [4]. It helps medical professionals to reduce the burden by providing automated and timely alerts.

This article presents patient-specific and patient-independent seizure prediction methods. An Ensemble learning strategy is proposed that incorporates three deep learning models: CNN1D, DenseNet1D, and CNN-LSTM1D model. This Ensemble learning mode facilitates an effective seizure prediction. Utilizing a single channel of EEG signals is proposed for effective seizure prediction, which makes the



**Fig. 1** Sample pattern of four states of EEG signals (CHB01, 22 channels).

proposed methodology suitable for designing lightweight and wearable seizure prediction devices. The subsequent sections of this article are organized as follows. Section 2 provides an overview of related works, while Sect. 3 describes the details of the proposed methods. Section 4 presents the experimental results including comparative studies with state-of-the-art works. Finally, Sect. 5 concludes the proposed methods.

## 2 Related Works

Over the past two decades, numerous research works have been developed focusing on seizure prediction. The majority of these methods are patient-specific seizure prediction [1, 2, 4]. A limited number of works have been developed for patient-independent seizure prediction [5, 6]. The EEG patterns are unique for each individual, which results in better performance for seizure-specific models compared to their independent counterparts. Before the era of deep learning, research in seizure prediction primarily relied on hand-crafted feature extractions and machine learning techniques, which demonstrated acceptable performances. The spatial features [1], spectral features [7], wavelet-transformed features [8], and fuzzy entropy features [9] were commonly used for seizure prediction, with SVM [7, 8], KNN [10], and MLP [11] being the prevalent machine learning techniques for feature classification.

The advent of deep learning has led to the application of various techniques for automatic feature extraction, capable of extracting features directly from raw data. CNN [2–4], DenseNet[12], LSTM [13], and Long-term recurrent convolutional networks [14] are among the commonly proposed deep learning techniques for seizure prediction. Researchers have employed various transformations such as continuous-wavelet transformation (CWT) [3], short-term Fourier transformation (SSTF) [15], discrete-wavelet transformation (DWT) [14], and Mel Frequency Cepstral Coefficients (MFCC) [5] to pre-process data for deep learning networks. Some researchers applied raw EEG signals without transformation for seizure prediction with acceptable accuracies [2, 4, 12]. Channel reduction is considered crucial for efficient seizure prediction. Several EEG channel reduction techniques have been proposed to enhance efficiency in seizure prediction [2, 4]. Seizure prediction using fewer channels will be more applicable for designing small-sized, power-efficient wearable seizure prediction devices. The performances of state-of-the-art works are not up to the mark using less number of EEG channels. Therefore, efficient patient-specific and patient-independent seizure prediction methods are proposed utilizing only one EEG channel.

### 3 Material and Methodology

In this work, the CHB-MIT standard database is employed for performance measures of our proposed patient-specific and patient-independent seizure prediction methods. This database is publicly available for research purposes [16]. It comprises EEG recordings from 23 pediatric epilepsy patients admitted at Children's Hospital, Boston. The EEG recordings were sampled at 256 Hz, following the International 10–20 system for electrode placement on the scalp. Each recording utilized at least 23 channels. Initially, 22 EEG channels are considered, as detailed in Table 1. These 22 channels are consistent across all EEG recordings of 23 patients. The database consists of 129 seizure recordings and 535 non-seizure recordings. The proposed methodology is structured into four components: the pre-processing of EEG signals, the learning technique, the channel selection technique, and the training and testing method.

#### 3.1 Pre-Processing of EEG Signals

The section outlines the pre-processing steps applied to the EEG signals. This work utilizes raw EEG signals for seizure prediction. An 8-second duration of EEG signal has proven sufficient for efficient seizure prediction [2, 4]. Consequently, an 8-second EEG signal duration is adopted as one sample in this work. A single EEG channel of 8-second duration encompasses 2048 signal intensity values ( $8 \times 256$ ), as the sampling frequency was 256 Hz during EEG recording. We have focused on utilizing only one EEG channel (Channel id: Ch12). An 8-second EEG data of one channel constitutes

**Table 1** Names of 22 unique channels of the CHB-MIT database

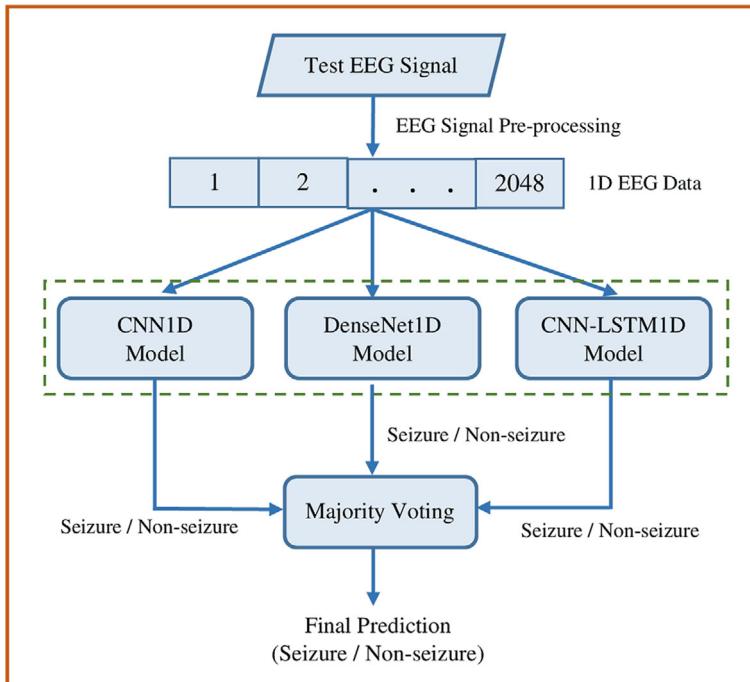
Channel id	Electrodes pair	Channel id	Electrodes pair
Ch01	FP1-F7	Ch12	P4-O2
Ch02	F7-T7	Ch13	FP2-F8
Ch03	T7-P7	Ch14	F8-T8
Ch04	P7-O1	Ch15	T8-P8
Ch05	FP1-F3	Ch16	P8-O2
Ch06	F3-C3	Ch17	FZ-CZ
Ch07	C3-P3	Ch18	CZ-PZ
Ch08	P3-O1	Ch19	P7-T7
Ch09	FP2-F4	Ch20	T7-FT9
Ch10	F4-C4	Ch21	FT9-FT10
Ch11	C4-P4	Ch22	FT10-T8

a 1-D array of size 2048. This data formation serves as one training sample for our proposed models.

### 3.2 Learning Technique for Seizure Prediction

Ensemble learning combines predictions from multiple individual models to improve predictive performance compared to any single model. Deep learning models like CNN [4], DenseNet [12], and CNN-LSTM [17] demonstrate commendable accuracy in seizure prediction. The 2D and 3D deep learning models involve a significant number of parameters. Therefore, 1D deep learning models are considered to design seizure prediction devices with reduced computational costs. The proposed ensemble learning approach combines three deep learning models: CNN1D, DenseNet1D, and CNN-LSTM1D. The final prediction relies on majority voting based on the predictions generated by these three models, as presented in Fig. 2.

In this work, the CNN1D model comprising six convolution layers and six max-pool layers is utilized for the extraction and reduction of features. Our proposed DenseNet1D model incorporates five dense blocks designed for various levels of



**Fig. 2** Proposed seizure prediction model using Ensemble learning.

feature extraction, each containing four convolution layers. The proposed CNN-LSTM1D model integrates six convolution layers, six max-pool layers, and ninety-six LSTM units for feature extraction and sequence prediction. During the convolution phase, input data  $I$  are convoluted with filter  $W$  to produce the output  $O$  according to (1). Here,  $O_j$  represents the convoluted output at location  $j$ , with  $W_k$  denoting the weight value of filter  $W$  at location  $k$ , where  $k$  ranges from  $-1$  to  $+1$ . For non-linearity in the output, the  $ReLU$  activation function is applied in each convolution layer, generating the same output for positive input values and zero output for negative input values, as expressed in (2). For feature classification, all three deep-learning models (CNN1D, DenseNet1D, and CNN-LSTM1D) are employed two fully connected layers. The number of nodes in the first fully connected layers is 128, 64, and 48 for CNN1D, DenseNet1D, and CNN-LSTM1D, respectively. In the second fully connected layer, two nodes with a  $Softmax$  activation function are employed across these models to derive probabilities for the two classes (inter-ictal and pre-ictal). The  $Softmax$  activation function is outlined in (3), yield output class probabilities. Here,  $p_i$  represents the probability of class  $i$ , and  $x_j$  denotes the input value of the activation function in class  $i$ , where  $j$  ranges from 1 to 2 for two classes (pre-ictal and interictal).

$$O_j = \sum_{k=-1}^{+1} I_{j-k} \times W_k \quad (1)$$

$$ReLU(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

$$Softmax(x_i) = p_i = \frac{e^{x_i}}{\sum_{j=1}^2 e^{x_j}} \quad (3)$$

### 3.3 Ranking of Channels Based on Channel Selection

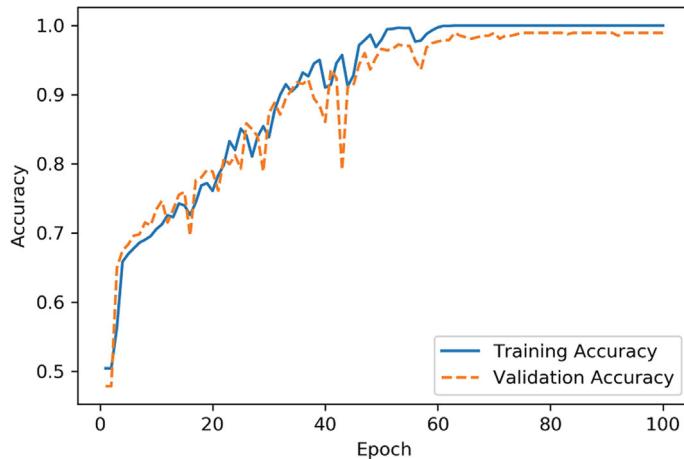
Channel selection is a crucial factor in the effectiveness of energy-efficient, wearable seizure prediction devices. Jana et al. have introduced an effective seizure prediction method utilizing only three EEG channels (Channel id: Ch11, Ch12, and Ch13) for optimal performance [4]. These channels correspond to five electrodes (C4, P4, O2, FP2, and F8). Designing seizure prediction devices with five EEG electrodes is not acceptable. Therefore, a single EEG channel is proposed consisting of only two electrodes. The varying performance of each channel for a specific deep-learning model indicates the different ranks of each channel at different times. The performance of each channel is evaluated across five rounds, and the channels are subsequently ranked based on their average performance. Finally, the channel (Channel id: Ch12) holding the 1st rank is deemed the most suitable for effective seizure prediction.

### 3.4 Training and Testing

The proposed models are evaluated in two approaches: patient-specific and patient-independent. Training and testing are conducted on patient-wise EEG data for the patient-specific method. Conversely, the patient-independent model utilizes a collective dataset, combining all patient data for training and testing. The model typically reaches saturation after 75 epochs, as illustrated in Fig. 3. Therefore, 100 epochs are considered for training.

During training, two non-seizure and two seizure recordings from each patient are considered, resulting in 46 seizure recordings and 46 non-seizure recordings for the patient-independent method. To create training samples, 10 min (600 s) of EEG data from each recording are employed, generating 593 training samples of 8 s each. For each patient, 1186 pre-ictal samples from two seizure recordings and 1186 inter-ictal samples from two non-seizure recordings are included. The proposed model involves 1305 pre-ictal samples from 87 seizures and 1275 samples from 85 non-seizure EEG recordings for testing. The binary cross entropy measurement is employed to calculate the total loss for this two-class classification problem, with  $y_i$  representing the actual class probability for class  $i$ , and  $n$  denoting the number of samples used for training, as mentioned in (4).

$$Loss = \frac{1}{n} \sum_{i=1}^n -(y_i \times \log(p_i) + (1 - y_i) \times \log(1 - p_i)) \quad (4)$$



**Fig. 3** Epoch-wise training and validation accuracy.

## 4 Results and Discussion

The evaluation of performance involves classification accuracy (*ACC*), precision (*PREC*), sensitivity (*SEN*), and specificity (*SPEC*). The performance evaluation utilizes the fivefold cross-validation technique. In the initial phase, individual channel performance was assessed through five rounds using the CNN-LSTM1D model. The performance is measured using the average values of accuracy, precision, sensitivity, and specificity, as outlined in (5). The results of round-wise performance, average performance, and the rank of each channel are detailed in Table 2. The ranking of each channel is based on the average performance. Remarkably, Ch12 emerged as the top-performing channel for seizure prediction.

$$\text{Performance} = \frac{\text{ACC} + \text{PREC} + \text{SEN} + \text{SPEC}}{4} \quad (5)$$

**Table 2** Round-wise performances and Rank of the individual channel.

Channel id	Round-wise performance					Average performance	Rank
	Round1	Round2	Round3	Round4	Round5		
Ch01	0.9806	0.9570	0.9831	0.9815	0.9797	0.9764	21
Ch02	0.9864	0.9786	0.9893	0.9676	0.9866	0.9817	16
Ch03	0.9914	0.9899	0.9887	0.9907	0.9885	0.9899	03
Ch04	0.9907	0.9926	0.9902	0.9868	0.9889	0.9899	02
Ch05	0.9831	0.9856	0.9828	0.9848	0.9813	0.9835	12
Ch06	0.9807	0.9870	0.9850	0.9775	0.9872	0.9835	13
Ch07	0.9867	0.9814	0.9848	0.9868	0.9866	0.9853	09
Ch08	0.9864	0.9883	0.9868	0.9916	0.9890	0.9884	05
Ch09	0.9832	0.9863	0.9750	0.9835	0.9870	0.9830	14
Ch10	0.9866	0.9672	0.9879	0.9877	0.9907	0.9840	10
Ch11	0.9864	0.9840	0.9880	0.9649	0.9874	0.9821	15
Ch12	0.9903	0.9915	0.9911	0.9891	0.9907	0.9905	<b>01</b>
Ch13	0.9831	0.9855	0.9846	0.9483	0.9838	0.9770	20
Ch14	0.9766	0.9868	0.9659	0.9782	0.9849	0.9785	19
Ch15	0.9670	0.9685	0.9864	0.9907	0.9879	0.9801	18
Ch16	0.9909	0.9910	0.9683	0.9925	0.9908	0.9867	07
Ch17	0.9865	0.9849	0.9862	0.9863	0.9858	0.9859	08
Ch18	0.9900	0.9829	0.9886	0.9841	0.9902	0.9871	06
Ch19	0.9883	0.9702	0.9688	0.9867	0.9871	0.9802	17
Ch20	0.9886	0.9884	0.9702	0.9815	0.9901	0.9838	11
Ch21	0.9841	0.9642	0.9637	0.9829	0.9749	0.9740	22
Ch22	0.9902	0.9907	0.9907	0.9883	0.9893	0.9898	04

**Table 3** Performance measures of proposed learning models.

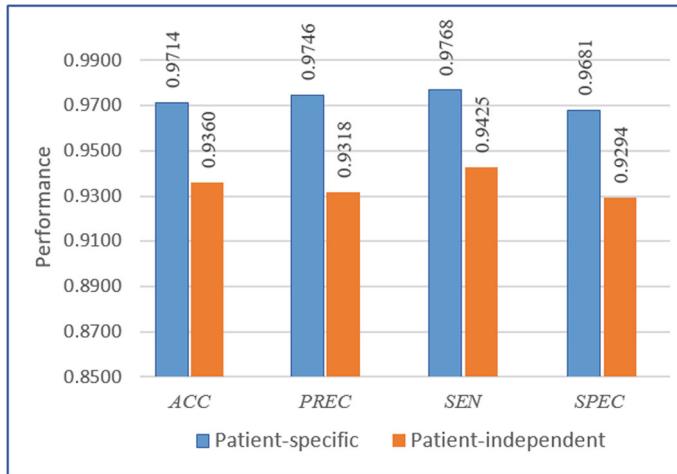
Learning technique	Patient-specific method				Patient-independent method			
	ACC	PREC	SEN	SPEC	ACC	PREC	SEN	SPEC
CNN1D	0.9507	0.9536	0.9551	0.9464	0.9186	0.9195	0.9195	0.9176
DenseNet1D	0.9587	0.9696	0.9514	0.9659	0.9244	0.9405	0.9080	0.9412
CNN-LSTM1D	0.9667	0.9674	0.9710	0.9623	0.9302	0.9310	0.9310	0.9294
Ensemble learning	0.9714	0.9746	0.9768	0.9681	0.9360	0.9318	0.9425	0.9294

Medical practitioners suggested using long-duration EEG signals for efficient seizure prediction. Jana et al. used fifteen consecutive samples with a duration of 8 s using majority voting to predict upcoming seizures [4]. Hence, fifteen consecutive samples are considered for majority voting in our proposed models. The performances of the CNN1D model, DenseNet1D model, CNN-LSTM1D model, and Ensemble learning model are evaluated. The Ensemble learning model provides superior results compared to individual models, as detailed in Table 3. Hence, the Ensemble learning model is proposed for patient-specific and patient-independent seizure prediction.

During testing, the proposed patient-specific model and patient-independent model using Ensemble learning provide excellent accuracy, sensitivity, and specificity, as mentioned in Fig. 4. The patient-specific model provides better performance compared to the patient-independent model. The proposed patient-specific prediction method is compared with state-of-the-art works that used the CHB-MIT database, as shown in Table 4. The proposed patient-independent prediction method is also compared with state-of-the-art works that used the same database, as shown in Table 5. The proposed two methods are the best among others. Our two proposed methods use only one EEG channel, which is appropriate for designing small-sized and low-power consumable prediction devices.

## 5 Conclusion and Future Research Directions

This article introduces effective patient-specific and patient-independent methods for predicting seizures. These methods predict seizures with ten-minute advance notice to take necessary precautions. These methods rely on predicting seizures using raw EEG signals. The proposed Ensemble learning approach employs leverages the majority voting technique to combine predictions from CNN1D, DenseNet1D, and CNN-LSTM1D models. It is observed that the Ensemble learning model provides better enhancing predictive performance than any individual model. A single EEG channel (Channel id: Ch12) is recommended for seizure prediction to optimize design considerations for the development of compact and low-power consumption seizure prediction devices. The patient-specific and patient-independent models provide



**Fig. 4** Comparative performances of patient-specific and patient-independent method using Ensemble learning.

**Table 4** Comparative performances of the proposed Ensemble learning model with state-of-the-art works (patient-specific).

Research work	Feature extraction	Learning technique	No. of channels	ACC	SEN	SPEC
Yao et al. [13]	Raw EEG data	Bi-LSTM	17	0.8780	0.8730	0.8830
Jana et al. [12]	Raw EEG data	CNN	06	0.9947	0.9783	0.9236
Guo et al. [14]	DWT	Easy-Ensemble	18	0.9262	0.9555	0.9257
Ryu et al. [18]	DWT	DenseNet and LSTM	22	0.9328	0.9292	0.9365
Shen et al. [19]	Tunable-Q wavelet transform	CNN	8	0.9757	0.9890	0.9790
Kapoor et al. [20]	Statistical features	Ensemble classifier	22	0.9661	0.9467	0.9136
Jana et al. [4]	Raw EEG data	CNN1D	03	0.9651	0.9655	0.9647
Proposed method	Raw EEG data	Ensemble learning	<b>01</b>	<b>0.9714</b>	<b>0.9768</b>	<b>0.9681</b>

sensitivity of 0.9768 and 0.9425, and specificity of 0.9681 and 0.9294, respectively. The proposed Ensemble learning provides better performances compared to the state-of-the-art works. It is observed that the patient-specific model provides better per-

**Table 5** Comparative performances of the proposed Ensemble learning model with state-of-the-art works (patient-independent).

Research work	Feature extraction	Learning technique	No. of channels	ACC	SEN	SPEC
Dissanayake et al. [5]	MFCC	CNN	23	0.8881	0.9345	0.8164
Dissanayake et al. [5]	MFCC	Siamese networks	23	0.9154	0.9245	0.8994
Halawa et al. [6]	DWT	CNN1D	18	0.9328	0.9292	0.9365
Proposed method	Raw EEG data	Ensemble learning	<b>01</b>	<b>0.9360</b>	<b>0.9425</b>	0.9294

formance compared to the patient-independent model. Future research directions involve improving the performance of the patient-independent model.

## References

1. Williamson JR, Bliss DW, Browne DW, Narayanan JT (2012) Seizure prediction using eeg spatiotemporal correlation structure. *Epilepsy Behav* 25(2):230–238
2. Jana R, Mukherjee I (2021) Deep learning based efficient epileptic seizure prediction with eeg channel optimization. *Biomed Signal Process Control* 68:102767
3. Khan H, Marcuse L, Fields M, Swann K, Yener B (2018) Focal onset seizure prediction using convolutional networks. *IEEE Trans Biomed Engin* 65(9):2109–2118
4. Jana R, Mukherjee I (2023) Efficient seizure prediction and eeg channel selection based on multi-objective optimization. *IEEE Access* 11:54112–54121
5. Dissanayake T, Fernando T, Denman S, Sridharan S, Fookes C (2021) Deep learning for patient-independent epileptic seizure prediction using scalp eeg signals. *IEEE Sens* 21(7):9377–9388
6. Halawa RI, Youssef SM, Elagamy MN (2022) An efficient hybrid model for patient-independent seizure prediction using deep learning. *Appl Sci* 12(11):5516
7. Zhang Z, Parhi KK (2016) Low-complexity seizure prediction from ieeg/seeg using spectral power and ratios of spectral power. *IEEE Trans Biomed Circuits Syst* 10(3):693–706
8. Janjarasjitt S (2017) Epileptic seizure classifications of single-channel scalp eeg data using wavelet-based features and SVM. *Med Biological Engin Comput* 55:1743–1761
9. Hussain W, Wang B, Niu Y, Gao Y, Wang X, Sun J (2019) Epileptic seizure detection with permutation fuzzy entropy using robust machine learning techniques. *IEEE Access* 7:182238–182258
10. Wang S, Chaovallitwongse WA, Wong S (2013) Online seizure prediction using an adaptive learning approach. *IEEE Trans Knowl Data Engin* 25(12):2854–2866
11. Ventura A, Franco JM, Ramos JP, Direito B, Dourado A (2009) Epileptic seizure prediction and the dimensionality reduction problem. Lecture notes in computer science, vol 5769. Springer, Heidelberg, pp 202–206
12. Jana R, Bhattacharyya S, Das S (2019) Epileptic seizure prediction from eeg signals using densenet. 2019 IEEE symposium series on computational intelligence. IEEE, China, pp 604–609

13. Yao X, Li X, Ye Q, Huang Y, Cheng Q, Zhang G (2021) A robust deep learning approach for automatic classification of seizures against non-seizures. *Biomed Signal Process Control* 64:102215
14. Guo Y, Jiang X, Tao L, Meng L, Dai C, Long X, Wan F, Zhang Y, Chen C (2022) Epileptic seizure detection by cascading isolation forest-based anomaly screening and easy-ensemble. *IEEE Trans Neural Syst Rehabilit Engin* 30:915–924
15. Shi S, Liu W (2024) B2-vit net: Broad vision transformer network with broad attention for seizure prediction. *IEEE Trans Neural Syst Rehabilit Engin* 32:178–188
16. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Stanley HE (1998) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, vol 101(23), pp 215–220. Circulation electronic pages
17. Wang Z, Zhou X (2022) Prediction of epileptic seizures based on cnn-lstm network. 2nd international conference on frontiers of electronics. Information and computation technologies. IEEE, China, pp 131–135
18. Ryu S, Joe I (2021) A hybrid densenet-lstm model for epileptic seizure prediction. *Appl Sci* 11(11):7661
19. Shen M, Wen P, Song B, Li Y (2023) Real-time epilepsy seizure detection based on eeg using tunable-q wavelet transform and convolutional neural network. *Biomed Signal Process Control* 82:104566
20. Kapoor B, Nagpal B, Jain PK, Gabralla LA (2023) Epileptic seizure prediction based on hybrid seek optimization tuned ensemble classifier using eeg signals. *Sensors* 23(1):423

# An Intuitive and Modular Framework for Enhanced Human-Machine Synergy



Mayank Kashyap, Apoorva Patel, Ashish Kumar, and Gurmeet Kaur

**Abstract** This paper proposes a novel Artificial Intelligence (AI) framework designed to enhance human-machine synergy through an intuitive and personalized approach. The framework integrates Large Action Models (LAMs), Large Language Models (LLMs), advanced computer vision, and a personalization engine to create an AI companion that fosters empathy and understanding. The proposed framework prioritizes ethical considerations, including accessibility, privacy, and security, and aims to comply with established ethical guidelines. The AI companion is expected to improve user engagement, task completion rates, and emotional intelligence through personalized and context-aware interactions. Potential applications include healthcare, education, entertainment, and social interaction. In healthcare, the AI companion could offer personalized treatment plans, remote patient monitoring, mental health support, and early disease detection. In education, it is expected to improve learning outcomes and motivation. This work lays the foundation for reshaping human-AI interaction, emphasizing ethical considerations and user needs. Future work will focus on implementation and evaluation against existing systems.

**Keywords** Multimodal interaction · Ethical AI design · Context-aware AI · Personalized user experience · Human-machine collaboration · Adaptive accessibility

---

M. Kashyap (✉) · A. Patel · A. Kumar · G. Kaur  
Department of Computer Science and Engineering, Chandigarh University, Mohali, Punjab, India  
e-mail: [contact@mayankkashyap.com](mailto:contact@mayankkashyap.com)

A. Patel  
e-mail: [contact@apoovapatel.com](mailto:contact@apoovapatel.com)

G. Kaur  
e-mail: [gurmeet.e15874@cumail.in](mailto:gurmeet.e15874@cumail.in)

## 1 Introduction

In an advancing field of Artificial Intelligence, it is an ever-increasing and pressing question to design more natural, intuitive, and beneficial human-machine interaction. In a world where AI penetrates our lives on an ever-growing scale, it is no longer plausible to regard it as a mere instrument to employ; rather, it takes up the role of a companion and collaborator opening up the new frontier filled with both promises and challenges. At the same time, the current reality demonstrates the state of human-machine interaction heavily short of the desired synergy with a wide gap between the two. The currently utilized AI systems are remarkable for their lack of adaptability to the user's individual needs, preferences, and personal features - in other words, they are used in the one-size-fits-all mode that allows no space for the human factor. Furthermore, the systems lack the natural, intuitive interfaces making the user adjust to the machine's ways of communication as opposed to an ideal image of AI adjusting to the user's preferred human style and modality. To bridge this gap, the new paradigm of AI system design is adopted where a more intuitive, personalized, and context-aware AI becomes a much-anticipated reality.

The main contributions of this work are as follows:

- The design of a modular and scalable AI framework that integrates cutting-edge technologies such as LLMs, LAMs, advanced computer vision, and a personalization engine to create an intuitive and personalized human-machine interaction experience.
- The incorporation of ethical considerations, including accessibility, privacy, and security, into the core design of the AI framework, ensuring that the technology is inclusive and protects user rights.
- The exploration of potential applications spanning healthcare, education, entertainment, and social interaction, demonstrating the versatility and broad impact of the proposed AI framework on society [10].

## 2 Related Work

The proposed AI framework builds upon and integrates insights from various cutting-edge AI technologies and products while setting itself apart through its emphasis on ethical considerations, accessibility, and the companion-based relationship between the AI and user.

Rabbit OS showcases the potential for multilingual capabilities, educational technology integration, and advanced AI features like voice interaction and action recording [1]. The Humane AI Pin offers a unique approach to organizing digital content with a centralized hub for photos, notes, and lists [2]. Ray-Ban Meta Smart Glasses introduce enhanced features and hands-free interaction through the integration of Meta AI, enabling immersive experiences and intuitive user interactions. Tab, the

world's first wearable AI, leverages advanced speech-to-text technology and sophisticated pattern recognition to provide an adaptive, personalized user experience.

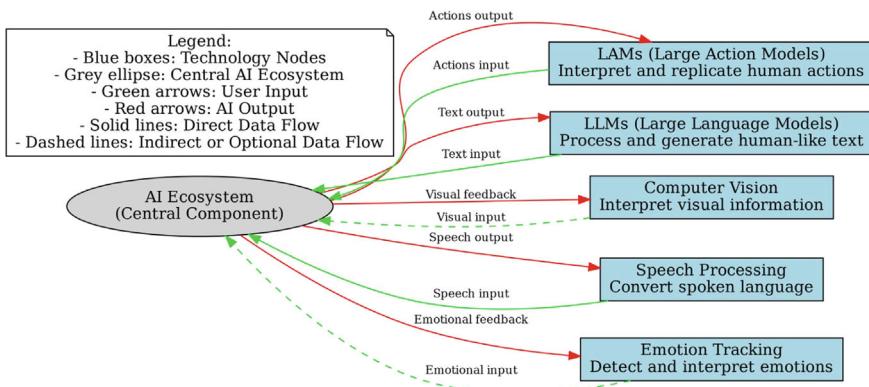
While these related works offer valuable insights, the proposed framework distinguishes itself through its strong emphasis on ethical considerations, universal design principles for accessibility, and the envisioned companion-based relationship between the AI and user. By integrating key innovations from these technologies while maintaining a strong ethical foundation, the proposed framework aims to create an intuitive, personalized AI companion that enhances human potential across diverse domains.

This comparative analysis situates the proposed framework within the broader context of AI innovation, highlighting opportunities to incorporate cutting-edge features while prioritizing user-centric design, accessibility, and ethical considerations. This synthesis of insights can guide future research and development efforts to realize the vision of an AI companion that empowers users and promotes the well-being and flourishing of humanity as a whole.

### 3 System Architecture and Technologies

The proposed AI framework aims to redefine human-machine synergy through a system architecture that integrates state-of-the-art technologies, harmonizing Large Action Models (LAMs), Large Language Models (LLMs), advanced computer vision, and innovative speech processing technologies to create a seamless, intuitive, and highly personalized interaction experience.

Figure 1 illustrates the modular system architecture of the proposed AI ecosystem, showcasing the integration of various cutting-edge technologies.



**Fig. 1** System architecture of the proposed AI ecosystem

### 3.1 Core Components

- **Large Action Models (LAMs):** LAMs interpret and replicate human actions within digital environments, combining neural and symbolic learning for efficient and explainable interaction processing.
- **Large Language Models (LLMs):** LLMs serve as the cornerstone of human-machine communication, employing Whisper for Speech to Text (STT), Eleven Labs API for Text to Speech (TTS), Large Language Model Meta AI 2 (LLaMA 2) for on-device processing of sensitive information, and GPT-4 API for advanced text generation tasks [3–6].
- **Advanced Computer Vision:** The framework integrates CogVLM for on-device image processing alongside TensorFlow, PyTorch, and Open Source Computer Vision Library (OpenCV), and uses the Canadian Institute For Advanced Research-10 (CIFAR-10) dataset for less sensitive computer vision tasks [7].

### 3.2 Integrating Technologies for a Seamless Experience

The modular design of the architecture facilitates scalability and adaptability, supporting real-time data processing on both the device and cloud. Visual data captured by cameras is processed by the computer vision models, and the resulting insights are fed into the LAMs and LLMs to understand context and generate appropriate responses, which are then converted to speech output using the Eleven Labs TTS API [4].

Personalization is achieved through the LAMs, which learn from user interactions over time to adapt the models to individual preferences and behaviors. The LLaMA 2 model processes sensitive personal data on-device to ensure privacy, while less sensitive tasks are handled by the cloud-based GPT-4 API [5, 6].

Verbal interactions are enabled through a combination of technologies. The Whisper LLM converts speech to text, which is then fed into the LLMs to generate a response. This textual response is converted back into natural-sounding speech using the Eleven Labs TTS API, facilitating smooth verbal communication between the user and the AI [3, 4].

To activate the system in a user-friendly manner, Porcupine is incorporated for wake-word detection, allowing users to initiate interactions with AI using simple verbal cues, enhancing accessibility and ease of use [8].

## 4 Personalisation and Data Management

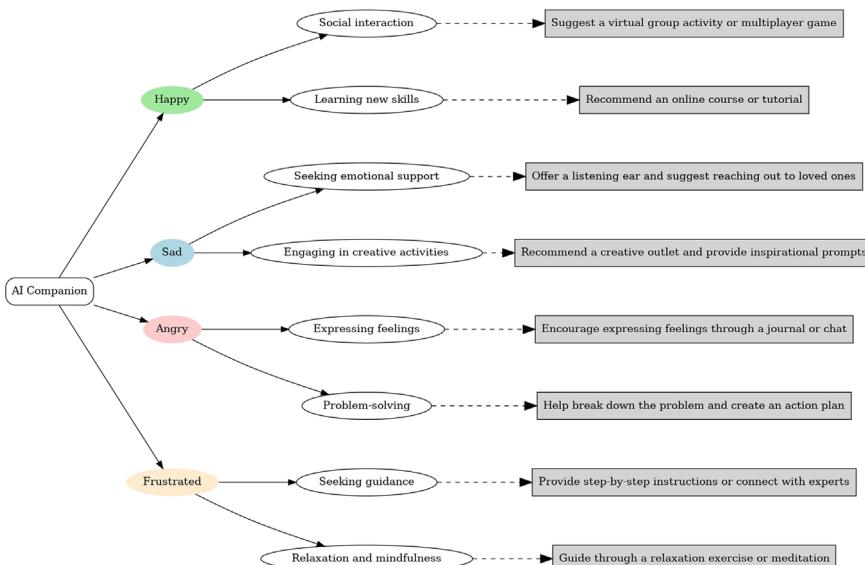
### 4.1 Personalisation Engine

The framework's personalization engine analyzes user data, preferences, and contextual information to create unique interaction profiles, adapting the functionality to meet individual needs. The personalization approach is based on user data analysis and machine learning techniques.

Figure 2 provides an overview of the personalization engine, illustrating how user data analysis and machine learning algorithms contribute to the creation of unique interaction profiles.

The following aspects form the foundation of the personalization approach:

- **User Data Analyzed:** The AI ecosystem collects and analyzes various types of user data, including demographic information, interaction history, preferences, emotional states, behavioral patterns, and feedback, to gain insights into user characteristics and tailor interactions accordingly.
- **Machine Learning Algorithms Employed:** The personalization engine leverages advanced machine learning algorithms, such as collaborative filtering, content-based filtering, reinforcement learning, clustering, and deep learning, to extract complex patterns and insights from user data and provide personalized experiences.
- **Creation and Adaptation of Interaction Profiles:** The AI ecosystem creates initial profiles based on user onboarding data and continuously updates and refines



**Fig. 2** Personalization engine overview

them based on collected data and analyzed patterns. Machine learning models are periodically retrained to adapt to evolving user behaviors and preferences, resulting in comprehensive interaction profiles that capture a wide range of user characteristics.

- **Examples of Personalized Interactions:** The AI ecosystem offers personalized interactions, including content recommendations based on user interests, emotional adaptation to provide appropriate support, contextual assistance tailored to specific situations, and interface customization to ensure accessibility and optimal usability for each individual.

## 4.2 Security and Privacy Framework

- **Encryption Methods:** The AI ecosystem employs robust encryption methods, such as end-to-end encryption, Advanced Encryption Standard (AES-256), Secure Sockets Layer (SSL), Transport Layer Security (TLS), and homomorphic encryption, to protect sensitive data during transit and at rest.
- **Local versus Cloud Processing Decision:** Sensitive data, such as personal information and biometric data, is processed locally on user devices to minimize security risks, while non-sensitive data and computationally intensive tasks are securely processed in the cloud. Users have the ability to define their own sensitivity thresholds and preferences for local versus cloud processing, and the AI ecosystem adheres to data minimization principles.
- **User Control Over Data:** The AI ecosystem provides users with granular privacy settings, allowing them to control what data is collected, shared, and processed. Clear opt-in and opt-out mechanisms are implemented, and users have the ability to review, update, and delete their personal data at any time. Data portability options are available, and privacy policies and terms of service are regularly updated and communicated to users.
- **Compliance with Data Protection Regulations:** The AI ecosystem is designed to comply with relevant data protection regulations, such as the General Data Protection Regulation (GDPR) for users in the European Union, the California Consumer Privacy Act (CCPA) for users in California, USA, and the Health Insurance Portability and Accountability Act (HIPAA) for handling sensitive health data.
- **Security Measures:** The AI ecosystem maintains a proactive approach to security by conducting regular security audits and penetration testing to identify and address vulnerabilities. Multi-factor authentication is implemented for user account access and sensitive actions, and secure coding practices and rigorous code reviews are enforced. Comprehensive incident response plans and procedures are established to promptly detect, investigate, and mitigate any security breaches or data incidents.

## 5 Advanced Sensors for Danger Detection and Health Monitoring

The integration of AI with medical sensors revolutionizes healthcare by enhancing health monitoring systems with off-body detection and near-body monitoring. AI algorithms improve disease prediction accuracy and clinical decision-making, enabling early detection, continuous monitoring, and personalized medicine [9].

The AI framework incorporates advanced sensors for danger detection and health monitoring. Environmental sensors detect potential hazards, providing proactive alerts to ensure user safety [9]. Wearable medical sensors enable personalized health insights, continuous monitoring, early issue detection, and recommendations for optimal well-being [9].

This holistic approach allows the AI framework to provide intuitive, personalized experiences while contributing to users' safety and health, offering a comprehensive solution that enhances various aspects of their lives.

## 6 AI as a Companion

The proposed AI Framework seeks to create a personalized and context-aware AI companion that adapts to the user's preferences, needs and emotional states. Through advanced personalization, such an AI companion alters its behavior and communication style and makes recommendations unique to each user and, thus, appears to "get" the user more. The multimodal interaction helps the AI to carry coherent conversations and bring the user-needed support while discussing matters he/she is intrigued about - as such, it provides much-needed emotional support and intellectual friendship. It remains essential to acknowledge the challenges and ethical implications behind companionship in AI, including additional reliance on artificial bonds, the authenticity of human-imitating AI behaviors, and a lack of true understanding of human emotion by AI. While it aims to simulate high-level personalized companionship and is contextually aware, it is imperative that the user is aware of the artificial nature of the conversation and is always encouraged to seek human conversations and support when in need.

## 7 Accessibility and Universal Design

### 7.1 *Importance of Accessibility and Inclusion*

The AI ecosystem prioritizes accessibility and inclusivity to ensure equal opportunities for all users, regardless of abilities or disabilities, to benefit from advancements in human-machine interaction.

## 7.2 *Implementing Universal Design Principles*

The seven principles of universal design are integrated into the AI ecosystem's development, making accessibility an integral part of the design and development process.

## 7.3 *Adaptability to Different Accessibility Needs*

The AI ecosystem adapts to various accessibility needs:

- Visual impairments: Audio descriptions of visual content.
- Hearing impairments: Real-time text transcriptions of audio.
- Mobility and dexterity limitations: Voice-activated controls and eye-tracking.
- Cognitive and learning disabilities: Simplified communication and visual aids.

## 7.4 *Continuous Improvement and User Feedback*

Accessibility is an ongoing process, with a commitment to continuous improvement through regular user feedback, dedicated user testing groups, and accessibility audits.

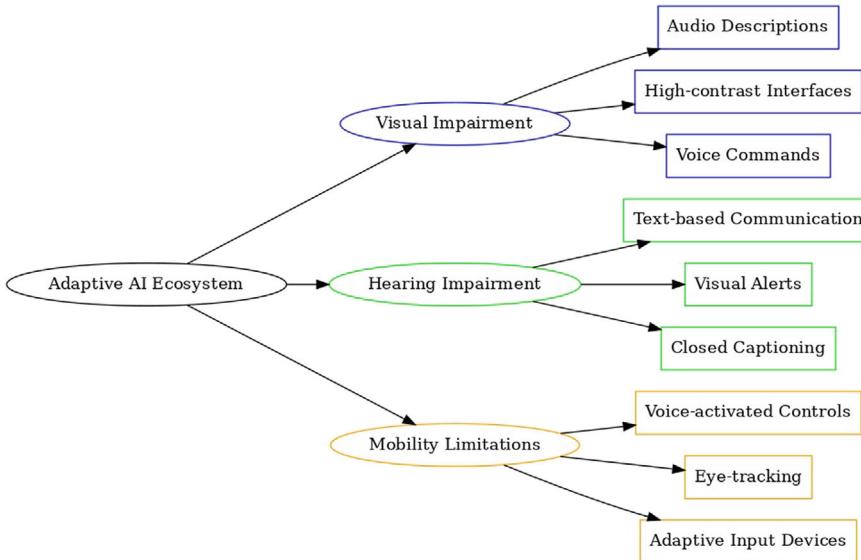
The AI ecosystem accommodates users with various impairments (Fig. 3), ensuring an inclusive experience that empowers users of all abilities to interact with and benefit from the technology.

## 8 Potential Applications

The AI ecosystem proposed in the presented work has every prospect to technologically reorganize numerous fields, given its unique capabilities of interaction and cooperation between humans and machines. In this paper, we address a series of critical usage areas in several fields to demonstrate the extent to which this system can affect societies.

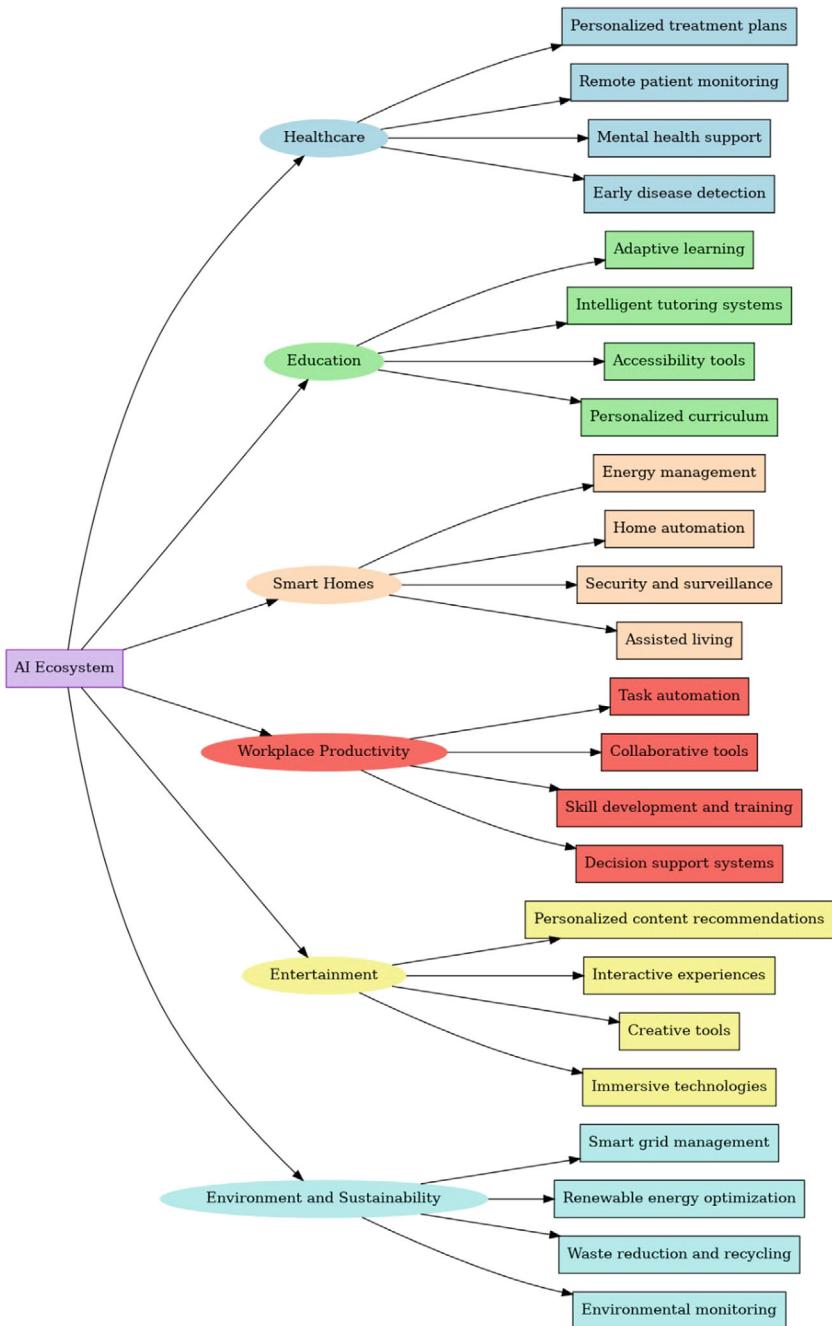
Figure 4 highlights the diverse potential applications of the proposed AI ecosystem across various domains, demonstrating its versatility and value.

- **Healthcare and Well-being:** By integrating AI with wearable and environmental sensors, healthcare providers can offer real-time, personalized monitoring and interventions for patients with chronic conditions or those undergoing rehabilitation. This proactive approach improves patient outcomes and reduces healthcare costs [9].



**Fig. 3** Accommodating users with different impairments

- **Education and Learning:** Personalized AI companions can transform the educational landscape by providing customized learning experiences. By understanding a student's learning style, pace, and preferences, AI can tailor educational content to maximize understanding and retention, support educators, and facilitate a more inclusive learning environment.
- **Smart Homes and Cities:** The AI ecosystem can enhance convenience, safety, and sustainability in smart homes and cities. Through intuitive human-machine interaction, residents can seamlessly control home appliances, lighting, and climate systems, leading to energy savings and improved comfort. The integration of environmental analysis and real-time monitoring can inform urban planning, optimize traffic flow, and enhance public safety.
- **Workplace and Productivity:** The system's ability to understand and replicate human actions can revolutionize workplace productivity tools. By automating routine tasks based on personalized user interactions, employees can focus on creative and complex problems, boosting efficiency and job satisfaction. The AI companion can also support mental well-being in the workplace by providing emotional support and mitigating feelings of isolation in remote work settings.
- **Entertainment and Social Interaction:** The AI companion concept extends into the realm of entertainment and social interaction, offering personalized content recommendations, engaging in meaningful conversations, and participating in interactive gaming. This enriches the user's leisure time and provides companionship to those who may feel isolated, promoting mental health and well-being [10].



**Fig. 4** Potential applications of the AI ecosystem

## 9 Methodology

- **Literature Review:** A comprehensive literature review of existing AI technologies and frameworks was conducted to identify gaps and limitations in current human-machine interaction approaches. Insights from related works, such as Rabbit OS, Humane AI Pin, Ray-Ban Meta Smart Glasses, and Tab, informed the design of the proposed framework [1, 2].
- **Framework Design:** Based on the literature review findings, a modular system architecture was conceptualized, integrating LAMs, LLMs, and advanced computer vision. The data flow and integration of multimodal components ensure seamless interaction between the user and the AI companion. A personalization engine leveraging machine learning algorithms and user data analysis was developed to achieve high levels of personalization. Advanced sensors for danger detection and health monitoring were incorporated to enhance the AI companion's capabilities [9].
- **Ethical Considerations:** Throughout the AI framework's development, accessibility and universal design principles were prioritized to ensure inclusivity and usability for individuals with diverse abilities. A robust security and privacy framework, including encryption methods and user data control, was designed to protect sensitive information. The framework ensures compliance with relevant data protection regulations, such as GDPR and CCPA, to uphold user privacy rights.
- **Iterative Refinement:** Based on the conceptual validation, an iterative refinement process was engaged to enhance the AI framework. Adjustments were made to the system architecture, data flow, and algorithm design to address identified challenges and incorporate improvements, allowing for continuous evolution and strengthening of the proposed AI framework.
- **Future Directions:** While the current research focuses on the conceptual design and methodology, further validation and refinement are acknowledged as important. Future work plans include developing a functional prototype of the proposed AI companion to conduct user studies, evaluate system effectiveness, and gather real-world feedback. Input from experts in AI, human-computer interaction, and ethics will be sought to further refine the framework and address potential challenges or ethical considerations.

## 10 Results and Conclusion

### 10.1 Results

The conceptual design of the proposed AI framework has been validated through a comprehensive literature review and the integration of state-of-the-art technologies. The framework's potential applications span various domains, including healthcare, education, entertainment, and social interaction, highlighting its versatility and value.

However, the research also identifies challenges and ethical considerations that must be addressed in future work, such as data privacy, transparency in AI decision-making, and the long-term effects of AI companionship on human behavior and well-being.

## 10.2 Conclusion

The proposed AI framework presents a novel approach to advancing human-machine synergy through an intuitive and personalized AI companion. By integrating LAMs, LLMs, advanced computer vision, and a personalization engine, the framework aims to foster empathy, understanding, and mutual growth between users and AI. The framework prioritizes ethical considerations, including accessibility, privacy, and security, through adherence to universal design principles and robust data protection measures.

The potential applications of this AI framework are vast, spanning healthcare, education, entertainment, and social interaction. However, realizing this vision requires addressing challenges related to data privacy, transparency in AI decision-making, and the ethical implications of AI companionship. Future research should focus on prototype development, real-world validation, and investigating the long-term effects of AI companions on human behavior and well-being. Collaboration across disciplines will be crucial for the responsible design and deployment of AI companions.

## References

1. Rabbit Research Team (2023) Learning human actions on computer applications. <https://rabbit.tech/research>
2. Mathew A (2023) Humane Ai pin - InnoGlove AI embrace. Int J Multidiscip Res (IJFMR) 5(6):1–5
3. Radford A, Kim JW, Xu T, Brockman G, McLeavey C, Sutskever I (2022) Robust speech recognition via large-scale weak supervision. In: 40th international conference on machine learning 2022. <https://cdn.openai.com/papers/whisper.pdf>
4. ElevenLabs Team (2023) ElevenLabs comes out of beta and releases eleven multilingual v2 - a foundational AI speech model for nearly 30 languages. ElevenLabs Blog, 22, 2023. <https://elevenlabs.io/blog/multilingualv2/>
5. Touvron H, Martin L, Stone K et al (2023) Llama 2: open foundation and fine-tuned chat models 19, 2023. arXiv preprint [arXiv:2307.09288](https://arxiv.org/abs/2307.09288)
6. OpenAI. GPT-4 technical report. arXiv (Cornell University), 2023. <https://doi.org/10.48550/arxiv.2303.08774>
7. Wang W, Lv Q, Yu W, Hong W, Qi J, Wang Y, Ji J, Yang Z, Zhao L, Song X, Xu J, Chen K, Xu B, Li J, Dong Y, Ding M, Tang J (2024) CogVLM: visual expert for pretrained language models. arXiv preprint [arXiv:2311.03079v2](https://arxiv.org/abs/2311.03079v2)
8. Chidhambararajan CR, Rangapur A, Chakkaravarthy S (2022) EfficientWord-Net: an open source hotword detection engine based on one-shot learning. J Inf Knowl Manag
9. Chen M, Cui D, Haick H, Tang N (2023) Artificial intelligence-based medical sensors for healthcare system. Adv Sens Res. <https://doi.org/10.1002/adsr.202300009>

10. Merrill K Jr, Kim J, Collins C (2022) AI companions for lonely individuals and the role of social presence. *Commun Res Rep* 39(2):93–103. <https://doi.org/10.1080/08824096.2022.2045929>
11. Agrawal D, Minocha S, Namasudra S, Kumar S (2021) Ensemble algorithm using transfer learning for sheep breed classification. In: IEEE 15th international symposium on applied computational intelligence and informatics (SACI). Timisoara, Romania, pp 199–204. <https://doi.org/10.1109/SACI51354.2021.9465609>
12. Gupta A, Namasudra S (2022) A novel technique for accelerating live migration in cloud computing. *Autom Softw Eng* 29:34. <https://doi.org/10.1007/s10515-022-00332-2>
13. Malviya S, Kumar P, Namasudra S (2022) Tiwary US (2022) Experience replay-based deep reinforcement learning for dialogue management optimisation. *ACM Trans Asian Low-Resour Lang Inf Process Just Accepted*. <https://doi.org/10.1145/3539223>

# Wavelet-Transformed K-NN Pipeline for EEG-Based Eye Blink Classification with Time Wrapping



N. Priyadharshini Jayadurga, M. Chandralekha, and Kashif Saleem

**Abstract** An innovative pipeline integrating wavelet transform, k-nearest neighbors (k-NN), and temporal wrapping approaches is presented in this work for the classification of eye blinks in electroencephalogram (EEG) recordings. The proposed process commences with raw EEG data preprocessing followed by wavelet treatment to extract significant time and frequency domain properties. These characteristics are subsequently utilized by a k-NN classifier to enhance the accuracy of eye blink recognition by using spatial relationships within EEG signals. Notably, the integration of a novel time wrapping technique deals with temporal dynamics, supplying the model with resilience to fluctuations in eye blink patterns across time. The pipeline's enhanced efficiency can be observed through validation using a customized EEG dataset, particularly when navigating the time complexity associated with eye blink events. Evaluations performed against current techniques demonstrate the efficacy of the proposed solution. The outcomes not only demonstrate the methodology's accuracy, but also its possible use in real-world settings featuring clinical diagnostics and human-computer interactions. This work conveys a strong basis for the discipline of EEG-based eye blink classification, having implications to enhance neuroscience and facilitating beneficial uses in healthcare and assistive technologies.

**Keywords** Temporal dynamics · Feature extraction · Data augmentation · Human-computer interaction · Cognitive functions · Neuroinformatics

---

M. Chandralekha and K. Saleem—These authors contributed equally to this work.

N. Priyadharshini Jayadurga · M. Chandralekha (✉)

Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Chennai, India  
e-mail: [m\\_chandralekha@ch.amrita.edu](mailto:m_chandralekha@ch.amrita.edu)

N. Priyadharshini Jayadurga

e-mail: [np\\_jayadurga@ch.students.amrita.edu](mailto:np_jayadurga@ch.students.amrita.edu)

K. Saleem

Department of Computer Sciences and Engineering, College of Applied Studies and Community Service, King Saud University, Riyadh, Saudi Arabia

e-mail: [ksaleem@ksu.edu.sa](mailto:ksaleem@ksu.edu.sa)

## 1 Introduction

The rapid confluence of neuroinformatics and signal processing has provided apparent methods for collecting important insights from Electroencephalogram (EEG) inputs. Reliable classification of eye blink instances is vital for grasping cognitive functions and human-computer interaction. In order to boost the efficacy of EEG-based eye blink classification, the present research proposes an original method that employs the strength of wavelet modification, the k-nearest neighbors (k-NN) algorithm, and temporal wrapping techniques. Despite the fundamental connection that exists between eye blinks and intellectual states, detecting eye blinks in EEG data is beneficial in a variety of contexts. Eye blinks can be classified as physiological phenomena, but their changes may indicate different diseases of a neurological nature such as epilepsy, cognitive load, or attention span. Detailing the purpose of the eye blink patterns would give a personalized clinical neurology intervention and therapy. It will disclose the huge field of possibility of human-computer interaction open in some cases. Such blink detectors potentially contribute to improved warning and give greater flexibility in brain-computer interfaces. Machine learning (ML) techniques enhance the extraction of patterns [1–3]. In addition, real-time blinking classification may serve as an early marker of cognitive fatigue in domains such as assistive technology and driver fatigue monitoring systems, consequently avoiding accidents while improving overall safety. This work emphasizes the significance of these kinds of initiatives by tackling the challenges related to EEG-based eye blink classification through an integrated pipeline. The current study contributes to the burgeoning body of knowledge in neuroinformatics by expanding the accuracy of blink detection in EEG signals while offering an avenue for applications in cognitive neuroscience, human-machine interaction, and healthcare.

Following the introduction, this paper is organized as Sect. 2 describing the related works involved in this domain, Sect. 3 discussed the proposed methodology, Sect. 4 states the results and discussion of the proposed work and Sect. 5 concludes the paper with the Conclusion.

## 2 Related Works

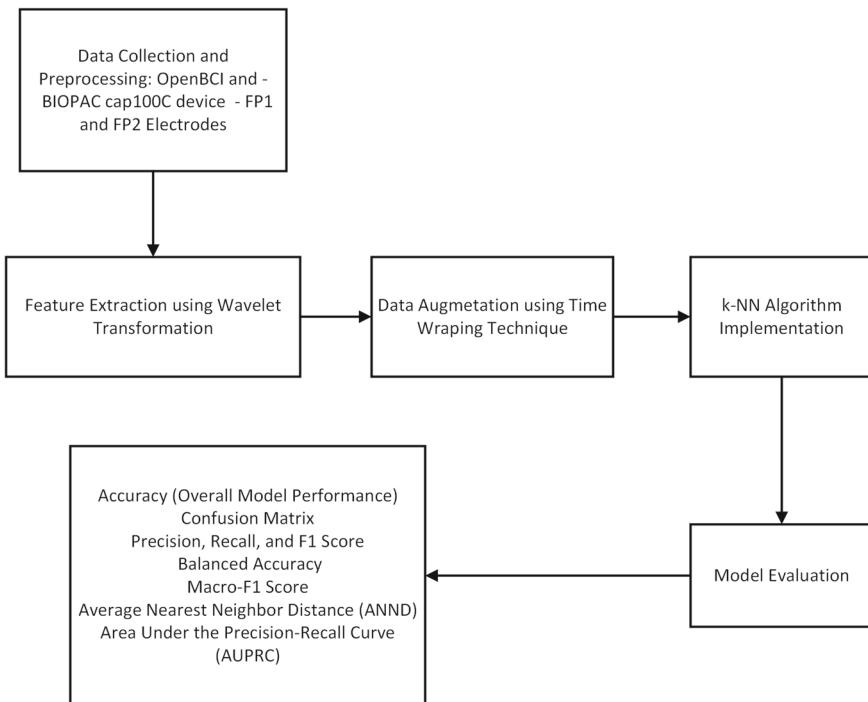
An article [4] combines polynomial-based feature extraction with kernel machines to improve EEG signal recognition for epileptic and visual conditions. Stronger discriminative features can be obtained via standard/kernel extension methods and polynomial transforms. On the Bonn-University database, sMLPNN notably surpasses LS-SVM exhibiting impressive predictivity, accuracy, and area under the receiver operating curve. The efficiency of kernel machine applications and comprehensive feature extraction in EEG-based diagnostics is demonstrated in this paper. The distinctive multi-level biometric identification method suggested in this paper [5] incorporates EEG and eye blinking EOG data. In an effort to improve EEG-

based authentication, the study explores fusion techniques at the feature and score levels. The effectiveness of combining eye blinking data in the proposed multi-level EEG biometric system is proven by evaluation on a dataset encompassing 31 subjects, indicating a notable improvement in accurate recognition and equal error rates. This study [6] emphasizes the transformational potential of Brain-Computer Interface (BCI) for empowering differently abled people through integrating Artificial Intelligence (AI) in the Internet of Things (IoT). The research extends previous techniques by employing a hybrid Deep Learning (DL) architecture to perform efficient EEG signal analysis. The constructed IoT-based BCI prototype was extensively validated using real-time EEG datasets and proved to be effective in real-world scenarios. This paper [7] investigates the utilization of co-registration of EEG and eye movement data to examine perceptual processes beneath free viewing conditions. It resolves problems with the self-paced nature of eye movements by providing a way to balance features across experiments. It underscores the importance of eliminating baseline selection issues by considering the effects of consecutive eye movements on EEG components and proposing strategies for EEG epoch segmentation. The work varies from typical stimulus-response paradigms and improves the study of visual perception in naturalistic situations by offering guidance to overcome methodological challenges. In the context of ubiquitous physiological sensors, this study [8] examines the relationship between eye gaze and electroencephalogram (EEG) data to determine boredom, a sense of being that has not received enough attention. By accomplishing a baseline study on physiological detection techniques and boredom, it employs an eye tracker, an EEG sensor, and a video stimulus to conduct an experiment. The results of this study reveal an intense connection between eye gazing and EEG data during periods of boredom in the participants ( $N = 13$ ). This work provides insightful information that promotes the development of systems which are aware of boredom. In a recent work [9], researchers devised an innovative EEG eye blink artifact identification algorithm for low spatial resolution data by integrating empirical mode decomposition (EMD) and Common Spatial Pattern (CSP) filtering with Particle Swarm Optimization (PSO) and Support Vector Machine. This method significantly excelled previous methods in detecting eye blink artifacts, as evidenced by its high accuracy and effectiveness when tested on EEG data from Zhejiang University School of Medicine's Children's Hospital. Rapid eye state classification [10] has become feasible utilizing a novel EEG-based algorithm that yields good accuracy in less than two seconds. This constitutes a significant improvement over previous methods that required 20 min. This approach delivers potential for enhanced real-time applications with more training data as it renders use of extensive deconstruction and machine learning algorithms. Children with autism were effectively classified in a study [11] which utilized EEG and eye-tracking data in a machine learning model with SVM. The approach, involving power spectrum analysis along with specific eye-tracking metrics on 97 children between the ages of 3 and 6, demonstrated potential as an ASD diagnostic tool. A recent paper [12] outlines a time-domain linear filtering technique which employs frontal electrodes and a multichannel Wiener filter to eradicate ocular artifacts from EEG, alleviating the need for a supplementary EOG sensor. This method performs superior for eye-blink cancellation and is

simpler to implement than traditional ICA techniques. With the goal of increasing user engagement, a recent study [13] reveals a hybrid brain-computer interface which combines natural gaze data from an eye tracker with motor imagery-based EEG classification. In a 2D cursor control assessment, this novel approach exhibited notable advancements, achieving over 80% accuracy and cutting down on job completion time.

### 3 Proposed Methodology

The proposed pipeline shown in Fig. 1 is composed of various steps like Data collection, Data Preprocessing, feature extraction, data augmentation, model implementation, and evaluation. These stages are discussed in detail in the following subsections.



**Fig. 1** Workflow of proposed pipeline

### 3.1 Feature Extraction

Wavelet transforms, particularly for orthonormal and biorthogonal wavelets with finite support, provide a new mathematical tool for dissecting continuous-time signals with potential applications in computer vision, signal coding, and various other areas [14]. Wavelet transformation is applied for optimal feature extraction in EEG-based eye blink classification. Applying this technique, a time-frequency representation corresponding to the raw EEG data is generated by splitting them into several different frequency components. From each resulting sub-band, statistical parameters such as mean, variance, skewness, and kurtosis are extracted. Time-domain features have been added, which include peak durations and amplitudes. The proposed approach provides an insight and description of EEG data that is able to handle the wide ranges of properties related to eye blink by considering both temporal dynamics and spectral features along with eye blink patterns. The model comes with many scales, and an addition to these is included to enhance the sensitivity level required for these delicate patterns, hence very reliable and efficient in its application for classification methods. The current approach derives its strength from a systematic process for feature extraction characterized by an array of mathematical expressions. The process begins with the application of wavelet transform (1), and it is followed by important statistical values being found, such as mean (Eq. (2)), variance (Eq. (3)), kurtosis (Eq. (5)) and skewness (Eq. (4)). Together with these calculations, they form the basis of the framework used in deriving features from EEG data.

$$W_x(a, b) = \frac{1}{\sqrt{a}} \int x(t) \psi^* \left( \frac{t-b}{a} \right) dt, \quad (1)$$

$$F_{\text{mean},i} = \frac{1}{N_i} \sum_{n=1}^{N_i} W_{xi}[n], \quad (2)$$

$$F_{\text{var},i} = \frac{1}{N_i} \sum_{n=1}^{N_i} (W_{xi}[n] - \mu_i)^2, \quad (3)$$

$$F_{\text{skew},i} = \frac{1}{N_i} \sum_{n=1}^{N_i} \left( \frac{W_{xi}[n] - \mu_i}{\sigma_i} \right)^3, \quad (4)$$

$$F_{\text{kurt},i} = \frac{1}{N_i} \sum_{n=1}^{N_i} \left( \frac{W_{xi}[n] - \mu_i}{\sigma_i} \right)^4 - 3 \quad (5)$$

where  $W_x(a, b)$  is the wavelet coefficient and  $F_i$  represents the feature vector for each sub-band  $i$ , giving mean, variance, skewness, and kurtosis. The combined feature vector can be given as Eq. 6.

$$F = [F_1, F_2, \dots, F_n, D_p, A_p] \quad (6)$$

where  $F$  is the final coupled feature vector,  $F_i$  stands for sub-band features,  $D_p$  stands for the duration of peak, and  $A_p$  stands for peak amplitude in the time space.

### 3.2 Data Augmentation

The framework outline for recognition of EEG-based eye blink signals has mentioned and classified time warping to be one of the important features in data enhancement through augmentation. The method correlates the temporal aspects of an EEG signal to a specified reference pattern, as depicted by [15]. The time-warping provides synthetic variations that affect the temporal dynamics of EEG eye blink-associated data through the change of temporal properties. The adoption of this augmentation approach advances the dataset to imitate various changes similar to the ones met in natural blinks but in time. The inclusion of time warping promotes the model's versatility and capacity to adapt to an extensive variety of temporal patterns by manipulating the temporal arrangement of the signals. Such an approach optimizes the model's robustness and efficacy by disclosing it to a wider range of temporal dynamics. This boosts the algorithm's ability to regulate fluctuations in eye blink patterns over time. The following is a brief summary of time warping in the context of EEG signal processing for eye blink classification:

Let  $x(t)$  resembles the original EEG signal as a function of time  $t$ , and  $x_{\text{ref}}(t)$  symbolizes the reference EEG pattern during eye blinks. The chronological wrapping function  $W$  aims to align  $x(t)$  with  $x_{\text{ref}}(t)$  as precisely as possible. The warped signal  $x_{\text{warped}}(t)$  is given in (7).

$$x_{\text{warped}}(t) = x(W(t)) \quad (7)$$

The warping function  $W(t)$  shifts the temporal scale of  $x(t)$  to match  $x_{\text{ref}}(t)$ . This non-linear function coincides with essential characteristics from  $x(t)$  to those in  $x_{\text{ref}}(t)$ .  $W(t)$  strives to minimize a distance or disparity measure between  $x_{\text{warped}}(t)$  and  $x_{\text{ref}}(t)$ , typically the Euclidean distance. Dynamic time warping (DTW) or similar algorithms can identify the best alignment between two time series.

### 3.3 Algorithm Implementation

The K-nearest neighbors (KNN) algorithm is renowned for its simplicity and efficacy in both classification and regression problems within the discipline of machine learning. The basic premise is that the classification of the unknown sample into one of the classes is based on a measure of proximity testing of how closely related the unknown is to the nearest K samples in the training data set. This classification identifies the majority category between the K nearest neighbors. k-NN therefore finds an application in eye blink recognition based on EEG, since it can make a good identification of the patterns from the spatial relationships within the feature space [16]. It estimates

the best number of neighbors for feature vectors input generated by combinations of wavelet processing and time-warping techniques in the cross-validation method. It optimizes its performance to respond to the spatial intricacies of the feature space of the algorithm. Some of the well-known uses of k-NN include recognition patterns in the classification of the EEG signals based on proximity to the neighboring locations in a multi-variant feature field. This real-time capability of quick changes in classification decisions for the use envisioned is very important, invaluable to any kind of instant reaction applications, for instance, human-computer interaction environments. No wonder, the k-NN turns out to be the simple, powerful, and effective classification technique easily bent towards the precise classification of EEG signals oscillating with the dynamics of eye blink. The following is the mathematical justification for the choice of the k-nearest neighbor algorithm in the identification of the eye blink pattern in an EEG.

To calculate the Euclidean distance  $d(\mathbf{x}, \mathbf{x}_i)$  within an unknown EEG signal sample feature vector  $\mathbf{x}$  and training set sample feature vectors  $\mathbf{x}_i$ , Eq. 8 is employed.

$$d(\mathbf{x}, \mathbf{x}_i) = \sqrt{\sum_{j=1}^n (x_j - x_{ij})^2} \quad (8)$$

where  $n$  is the total amount of features,  $x_j$  is the  $j$ -th feature of  $\mathbf{x}$ , and  $x_{ij}$  is the  $j$ -th feature of  $\mathbf{x}_i$ .

After determining K nearest neighbors, the sample that is unidentified  $\mathbf{x}$  is classified based on its most frequent class between these neighbors by using Eq. 9.

$$\hat{C} = \text{mode}\{C_1, C_2, \dots, C_K\} \quad (9)$$

where  $C_k$  is the class of the  $k$ -th nearest neighbor, and  $\hat{C}$  is the predicted class for  $\mathbf{x}$ .

## 4 Performance Analysis

### 4.1 Dataset Details

This study utilized publicly available EEG data [17–19] from the internet to investigate EEG-based eye blink classification. The OpenBCI Device and BIOPAC Cap100C were employed for recording 20 individuals, with particular focus on frontal electrodes Fp1 and Fp2. Each participant deliberately performed one eye-blink, with additional blinks forced on by outside stimuli. In a single session, each subject stipulated around twenty-five blinks. The dataset was meticulously annotated by hand with eye blink events which were synchronised with EEG signals from video channels. Preprocessing techniques comprised baseline correction and artifact removal. This readily available data supports the background that was addressed in

earlier exchanges and is a beneficial input for the pipeline that is being proposed for EEG-based eye blink classification.

## 4.2 *Experimental Setup*

This study adopted a robust experimental strategy for classifying EEG signals associated with eye blink occurrences. During analysis, the corresponding ground truth labels and the unstructured EEG data in a DataFrame were used. A time-warping function has been added to the dataset to enhance it by providing variations in the temporal dynamics of the EEG signals. Wavelet transformation was important to feature extraction, with the Daubechies 1 wavelet with a level of 3 utilized to extract relevant features. After this, the training data was exposed to time warping, and then additional feature extraction employing wavelet transformation yielded a substantial feature matrix. The dataset was conventionally resized and reshaped in accordance to the analysis's input criteria. The study concentrated on traditional machine learning approaches, particularly the use of a K-NN classifier with 5 neighbors. This classifier was trained using scaled wavelet-transformed training data and then used to predict scaled test data. For the K-NN model, evaluation metrics like confusion matrices, accuracy scores, and classification reports were determined. This revised experimental setup, based on advanced signal processing and conventional machine learning, provides an adequate basis for studying EEG-based eye blink classification strategies.

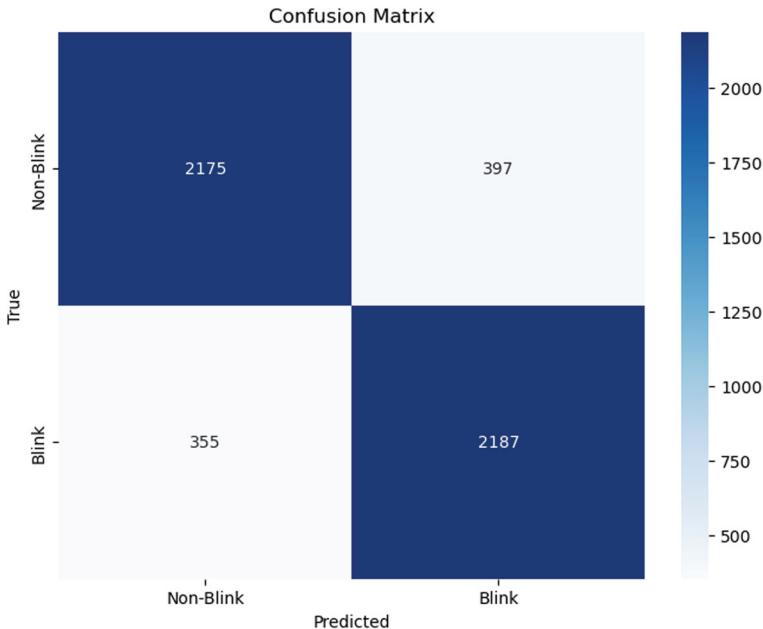
## 4.3 *Results and Discussions*

### Accuracy (Overall Model Performance)

The frequency that the model effectively predicts outcomes is determined by its accuracy. It represents the proportion of accurate predictions among all of the model's estimates. The reported accuracy of 85.29% indicates the proportion of correctly classified instances in the total. This indicates to an overall performance which is satisfactory, capturing an elevated number of accurate predictions in the blink and no-blink classes.

### Confusion Matrix (Blink vs. No Blink)

The model's predictions are split down into cases where the confusion matrix adequately identified events (true positives and true negatives) and cases where it misinterpreted events (false positives and false negatives). The confusion matrix provides an extensive summary of the model's predictions. It is relatively simple to evaluate



**Fig. 2** Confusion matrix of proposed pipeline

the model's strengths as well as potential areas for improvement when one is aware of the true positives, true negatives, false positives, and false negatives. For an additional assessment, these elements need to be analyzed. The confusion matrix for the problem at hand is shown in Fig. 2.

#### Precision, Recall, and F1 Score

Precision assesses the extent to which optimistic predictions come true. Recall evaluates the degree to which a model comprehends each and every occurrence of a positive class. The F1 score establishes a single metric that takes into account both false positives and false negatives by achieving a balance across precision and recall. Collectively, precision, recall, and the blink class's F1 score (85%, 86%, and 85%, respectively) illustrate the model's ability to recognize eye blinks with consistency. A balance between precision and recall is critical, and these metrics show superior results in detecting positive results while restricting false positives.

### Balanced Accuracy

An adjusted version of accuracy that takes imbalances in class distribution into account is referred to as balanced accuracy. By guaranteeing the model's performance is assessed equitably across classes, it eliminates biases caused by unequal class representation. The balanced accuracy of 85.30%, allowing for class imbalance, is more indicative of the model's overall performance. This statistic is one way of controlling biases in the result from the imbalanced datasets, which show the constant accuracy of the model across classes: "blink" and "no blink."

### Macro-F1 Score

The average of the F1 scores calculated sequentially for each class is termed the macro-F1 score. It serves as a broad indicator of how well the model distinguishes across various classes. An overall macro-F1 score estimated as 85.30% in the proposed method provides a comprehensive measurement of the effectiveness of the model without downplaying the importance of higher individual F1 scores from each category. This measure couples with balanced accuracy, suggesting a consistent well-rounded and uniform performance across all classes.

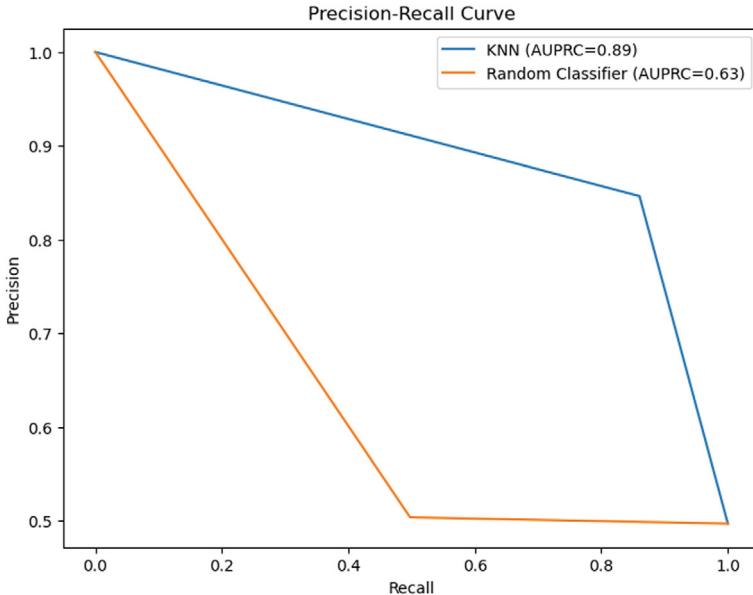
### Average Nearest Neighbor Distance (ANND)

The mean distance in the feature space between data points is obtained by ANND. Instances pertaining to the same class are believed to be closer to one another whenever the ANND is smaller, demonstrating effective separation in the feature space. The average distance in feature space between data points is determined by the average nearest neighbor distance, whose value is 0.00019418. A smaller ANND indicates that instances in the same class appear more closely associated, indicating successful feature space separation and robust clustering.

### Area Under the Precision-Recall Curve (AUPRC)

When a model exhibits a disparity between positive and negative instances, the model's ability to strike an equilibrium between Precision and recall is evaluated using the Area Under the Precision-Recall Curve (AUPRC). The KNN model's reported Area Under the Precision-Recall Curve (AUPRC) of 0.888 implies that it performs well while analyzing unbalanced data. This statistic is especially useful when evaluating models in cases where one class is more prevalent than the other, as it accentuates the model's ability to effectively balance precision and recall. The AUPRC of k-NN algorithm in comparison to a random classifier is depicted in Fig. 3.

In combination, these metrics demonstrate that the model for EEG-based eye blink classification is precise, balanced across classes, and resilient when confronted with

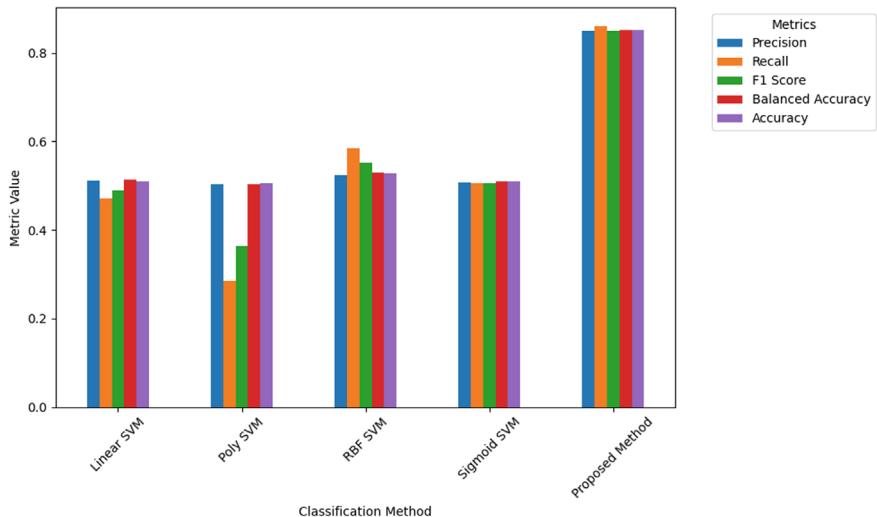


**Fig. 3** AUPRC of proposed pipeline

unbalanced datasets. The extensive evaluation metrics provide substantial insight into multiple facets of the model's performance, permitting an exhaustive understanding of its advantages and disadvantages.

### Comparison of Results

The proposed pipeline was compared with the Support Vector Machine algorithm using various kernels, as it was predominantly used for similar tasks. The graph 4 compares key performance metrics (Precision, Recall, F1 Score, Balanced Accuracy, and Accuracy) of various SVM kernels and the proposed method. While accuracy values for SVM kernels are not provided, the proposed method demonstrates superior performance across all metrics, indicating its effectiveness compared to traditional SVM approaches. The following Table 1 presents the comparison of performance metrics between the proposed pipeline and various SVM kernels.



**Fig. 4** Comparative performance of SVM kernels and proposed EEG classification method

**Table 1** Comparison of performance metrics between the proposed method and SVM kernels

Method	Precision	Recall	F1 Score	Balanced accuracy
Linear SVM	0.5115	0.4709	0.4904	0.5132
Poly SVM	0.5045	0.2844	0.3638	0.5042
RBF SVM	0.5241	0.5854	0.5531	0.5300
Sigmoid SVM	0.5073	0.5051	0.5062	0.5101
Proposed Method	0.85	0.86	0.85	0.85

## 5 Conclusion

In a nutshell, the pipeline for characterizing eye blinks based on electroencephalogram (EEG) exhibited impressive outcomes, with an accuracy of 85.29% and strong performance metrics across multiple evaluation standards. Metrics like precision, recall, and F1 score revealed the extent to which the model identified positive events, emphasizing its potential for practical use. Future developments involve examining more complex deep learning architectures, incorporating ensemble methods, transferring the system into practice in real time, such as online learning tactics, and assessing findings across several datasets. These endeavors are aimed at strengthening the model's effectiveness with regard to precision, flexibility, and generalizability, positioning it at the forefront of EEG-based categorization systems for sophisticated neuroscientific studies and dynamic scenarios featuring human-machine interaction.

## References

1. Mageshwari G, Chandrakha M, Chaudhary D (2023) Underwater image re-enhancement with blend of simplest colour balance and contrast limited adaptive histogram equalization algorithm. In: 2023 international conference on advancement in computation and computer technologies (InCACCT), pp 501–508. <https://doi.org/10.1109/InCACCT57535.2023.10141807>
2. Durga Bhavani K, DRN (2020) K-means clustering using nature-inspired optimization algorithms-a comparative survey. *Int J Adv Sci Technol* 29(6s):2466–2472
3. Aswiga RV, Karpagam M, Chandrakha M, Kumar CS, Selvi M, Deena S (2023) An automatic detection and classification of diabetes mellitus using cnn. *Soft Comput* 27(10):6869–6875. <https://doi.org/10.1007/s00500-023-08122-y>
4. Djoufack Nkengfack LC, Tchiotsop D, Atangana R, Tchinda BS, Louis-Door V, Wolf D (2021) A comparison study of polynomial-based pca, kPCA, lda and gda feature extraction methods for epileptic and eye states eeg signals detection using kernel machines. *Inf Med Unlocked* 26:100721. <https://doi.org/10.1016/j.imu.2021.100721>
5. Abo-Zahhad M, Ahmed SM, Abbas SN (2016) A new multi-level approach to EEG based human authentication using eye blinking. *Pattern Recognit Lett* 82:216–225. <https://doi.org/10.1016/j.patrec.2015.07.034>
6. Medhi K, Hoque N, Dutta SK, Hussain MI (2022) An efficient eeg signal classification technique for brain-computer interface using hybrid deep learning. *Biomed Signal Process Control* 78:104005. <https://doi.org/10.1016/j.bspc.2022.104005>
7. Nikolaev AR, Meghanathan RN, van Leeuwen C (2016) Combining eeg and eye movement recording in free viewing: pitfalls and possibilities. *Brain and Cognition* 107:55–83. <https://doi.org/10.1016/j.bandc.2016.06.004>
8. Kim J, Seo J, Laine TH (2018) Detecting boredom from eye gaze and eeg. *Biomed Signal Process Control* 46:302–313. <https://doi.org/10.1016/j.bspc.2018.05.034>
9. Wang M, Cui X, Wang T, Jiang T, Gao F, Cao J (2023) Eye blink artifact detection based on multi-dimensional eeg feature fusion and optimization. *Biomed Signal Process Control* 83:104657. <https://doi.org/10.1016/j.bspc.2023.104657>
10. Saghafi A, Tsokos CP, Goudarzi M, Farhidzadeh H (2017) Random eye state change detection in real-time using eeg signals. *Expert Syst Appl* 72:42–48. <https://doi.org/10.1016/j.eswa.2016.12.010>
11. Kang J, Han X, Song J, Niu Z, Li X (2020) The identification of children with autism spectrum disorder by svm approach on eeg and eye-tracking data. *Comput Biol Med* 120:103722. <https://doi.org/10.1016/j.combiom.2020.103722>
12. Borowicz A (2018) Using a multichannel wiener filter to remove eye-blink artifacts from eeg data. *Biomed Signal Process Control* 45:246–255. <https://doi.org/10.1016/j.bspc.2018.05.012>
13. Dong X, Wang H, Chen Z, Shi BE (2015) Hybrid brain computer interface via bayesian integration of eeg and eye gaze. In: 2015 7th international IEEE/EMBS conference on neural engineering (NER), pp 150–153. <https://doi.org/10.1109/NER.2015.7146582>
14. Akansu AN, Haddad RA (2001) Chapter 6 - wavelet transform. In: Akansu AN, Haddad RA (eds) Multiresolution signal decomposition, 2nd edn. Academic, San Diego, pp 391–442. <https://doi.org/10.1016/B978-012047141-6/50006-9>. <https://www.sciencedirect.com/science/article/pii/B9780120471416500069>
15. Iwana BK, Uchida S (2020) Time series data augmentation for neural networks by time warping with a discriminative teacher
16. Shyam KP, Ramya V, Nadiya S, Parashar A, Gideon DA (2023) Chapter 15 - systems biology approaches to unveiling the expression of phospholipases in various types of cancer-transcriptomics and protein-protein interaction networks. In: Chakraborti S (ed) Phospholipases in physiology and pathology. Academic, pp 271–307. <https://doi.org/10.1016/B978-0-443-15177-4.00016-9> . <https://www.sciencedirect.com/science/article/pii/B9780443151774000169>

17. Agarwal M, Sivakumar R (2019) Blink: A fully automated unsupervised algorithm for eye-blink detection in eeg signals. In: 2019 57th annual allerton conference on communication, control, and computing (Allerton), pp 1113–1121. <https://doi.org/10.1109/ALLERTON.2019.8919795>
18. Agarwal M, Sivakumar R (2020) Charge for a whole day: extending battery life for bci wearables using a lightweight wake-up command. In: Proceedings of the 2020 CHI conference on human factors in computing systems, CHI '20. Association for Computing Machinery, New York, pp 1–14. <https://doi.org/10.1145/3313831.3376738>
19. Gupta E, Agarwal M, Sivakumar R (2020) Blink to get in: biometric authentication for mobile devices using eeg signals. In: ICC 2020 - 2020 IEEE international conference on communications (ICC), pp 1–6 (2020). <https://doi.org/10.1109/ICC40277.2020.9148741>

# Enabling Cursor Control Through Eye Movement Using Hidden Markov Model



G. Tanusha , P. Havarbhavi , and K. Ashwini

**Abstract** First and foremost, we present a novel scheme for controlling the cursor based on eye movement. Other strategies rely on physical items like touchpads or a mouse and place strict requirements on people with motor handicaps. Our solution connects digital interaction with human visual attention through eye-tracking technology. We examine the nuanced gaze patterns that develop, allowing for cursor control while also developing a sense of interaction context. Gaze data is transformed by our method into the motion of a cursor. It is highly accurate and adjustable. Cursor motions are generated based on the gaze movement pattern, making the interface more dynamically accurate and more tailored to individual behaviors. The context-aware interface maximizes efficiency through its adaptability as it learns and adopts the behaviors. In addition, our research introduces a novel scheme through the use of software tools to offer an eye-controlled mouse system. This system accurately recognizes facial features and interprets eye movements through image processing techniques. The system enables intuitive cursor control by interpreting observable eye movements. Transforming eye motions into a cursor's actions, it enhances user accessibility and hands-free interaction; in other words, its performance is captured in the experimental analysis.

**Keywords** Eye-controlled cursor interface · Visual tracking technology · Pattern analysis · Software tools · User accessibility · User-centric design · Adaptive systems

---

G. Tanusha · P. Havarbhavi · K. Ashwini

Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Chennai, India

e-mail: [tanushag1408@gmail.com](mailto:tanushag1408@gmail.com)

P. Havarbhavi

e-mail: [havarbhavi1904@gmail.com](mailto:havarbhavi1904@gmail.com)

K. Ashwini

e-mail: [k\\_ashwini@ch.amrita.edu](mailto:k_ashwini@ch.amrita.edu)

## 1 Introduction

It is essential to achieve seamless human–computer connection in the modern technological world. Although mice and touchpads are common input devices, people with motor difficulties may find them difficult to use. Accessibility problems are exacerbated since people who are unable to physically operate these gadgets are frequently excluded from digital activities. Researchers and engineers are looking into creative ways to close the gap between technology and human capabilities to overcome these constraints. With the use of eye-tracking technology, digital gadgets may be operated hands-free thanks to intuitive interfaces that respond to visual signals. The development of eye-controlled cursor systems, which make use of patterns of eye movement to enable smooth navigation of digital interfaces and accurate cursor manipulation, is a noteworthy breakthrough in this field.

Using concepts from pattern analysis and computer vision, this research project investigates the creation of a novel eye-controlled cursor system. The foundation of our approach is the use of analytical methods to translate complex eye movement patterns into precise cursor motions. Our goal is to create a robust and adaptable control system that can appropriately accommodate a variety of user behaviors and preferences by training with observed eye movement data. We hope to go above the limitations of conventional cursor control methods and improve accessibility for people with motor disabilities by fusing cutting-edge methods with assistive technology resources. Our study seeks to pave the way toward a more comprehensive and adaptable human–computer interaction, empowering users of diverse abilities to interact with digital technology autonomously and effortlessly.

### 1.1 *Eye-Controlled Cursor System*

A novel cursor control system has been developed using statistical models like the Hidden Markov Model and eye movement tracking. This innovation allows for hands-free computer interaction, significantly enhancing accessibility for individuals with limited motor function.

### 1.2 *Data Processing*

The system employs advanced pattern recognition (pupil analysis) and image analysis techniques to convert complex eye movements into precise cursor control on the screen.

### ***1.3 Adaptive Interface***

The interface is designed to be context-aware, improving accuracy and efficiency by learning and adapting to each user's unique interaction patterns over time.

### ***1.4 Practical Testing***

Extensive research and trials have been conducted to assess the eye-controlled system's real-world effectiveness and usability, focusing on future applications and improvements.

The paper provides a thorough summary of our work on employing a Hidden Markov Model to enable cursor control using eye movement. It begins with an Introduction that summarizes the purpose, goals, and significance of the study. Related Work follows, reviewing and highlighting relevant literature and pointing out any gaps. The design, development, collecting data, and image processing methods are covered comprehensively in the Methodology section. The combination of hardware and software is described by system architecture. Testing procedures and metrics are provided in Experimental Setup and Evaluation, and results are evaluated and compared with current solutions in Results and Discussion. The document ends with a Conclusion that provides a summary of the main conclusions and their consequences as well as references. Our thorough examination and testing determine the approach's feasibility and effectiveness, providing the foundation for its use. We see a time where eye-controlled pointer technologies allow people to interact with technology without limitations.

## **2 Related work**

Information on several strategies for enhancing eye movement-based HCI for people with impairments may be found in the mentioned studies. Sivasangari et al. [1] and Ganga et al. [2] replaced conventional input devices like mice and keyboards with eye motions using a mix of Raspberry Pi pupil detection and webcam-based eye tracking. Based on eye movements, hands-free cursor control methods were developed by Mangaiyarkarasi and Geetha [3] and Dhanasekar et al. [4]. These systems employ machine learning and image processing techniques to properly track and analyze eye movements. Human-computer interaction systems that detect eye movements to enable hands-free cursor control and device communication through eye gaze were proposed by Narahari et al. [5] and Chandra et al. [6] using approaches like eye-aspect ratio and nonverbal communication.

Emphasis is placed on the challenges posed by conventional input devices, with particular attention to musculoskeletal disorders brought on by prolonged computer

usage [2]. A number of the proposed solutions include external equipment, such as infrared-based eye-tracking systems or head-mounted devices [3, 5]. For average individuals, however, implementing these solutions could be costly and challenging [4]. In contrast, the approach proposed by Sivasangari et al. [1] makes use of a simple camera that does not require any infrared technology and offers a workable and reasonably priced replacement.

A hands-free cursor control system based on eye movements is provided by Dhanasekar et al. [4] as an example of how machine learning algorithms may enhance accessibility. Furthermore, Narahari et al. [5] and Ganga et al. [2] study eye-based interaction systems, emphasizing how important accurate eye movement tracking and interpretation is for managing cursor control. In addition to highlighting the importance of nonverbal engagement strategies, Chandra et al. [6] provide a method that assists physically challenged individuals through eye gazing. When taken as a whole, these studies demonstrate the significance of eye movement-based engagement techniques for enhancing the accessibility and independence of individuals with disabilities. Hidden Markov Models (HMMs) are a useful tool for improving the effectiveness and adaptability of cursor control systems for individuals with impairments, which provide an additional level of complexity to the study and prediction of eye movements.

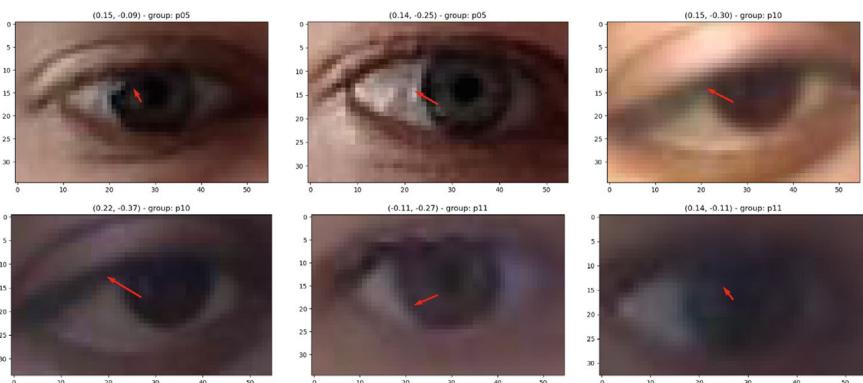
The majority of the approaches discussed for eye tracking and cursor control rely on conventional image processing techniques or machine learning algorithms; however, the use of Hidden Markov Models (HMMs) offers a fresh perspective in this field. Unlike previous approaches that primarily focus on directly mapping eye movements to cursor positions, HMM-based systems can uncover underlying patterns and states related to gaze behaviors. With a more flexible and dependable base for cursor control, this capability might lead to improved accuracy and usability in human-computer interaction.

In the realm of research leveraging Hidden Markov Models, the proposed method for detecting fraudulent activities in electronic auctions by employing HMM in conjunction with the X-means clustering algorithm [7]. This advancement parallels the application of HMM in enabling cursor control through eye movement, showcasing the versatility of HMM across diverse domains, from fraud detection to assistive technology development. The research by Havarbhavi et al. proposes a novel method for controlling the cursor using eye movements. By utilizing visual tracking technology instead of physical devices like mice, this approach enables an effortless interaction experience, especially beneficial for those with mobility impairments. Its flexibility and ability to adapt to individual user behaviors represent a notable advancement in seamless human-computer integration, signifying a crucial intersection where human capabilities and technological solutions converge within the domain of cursor control interfaces [8].

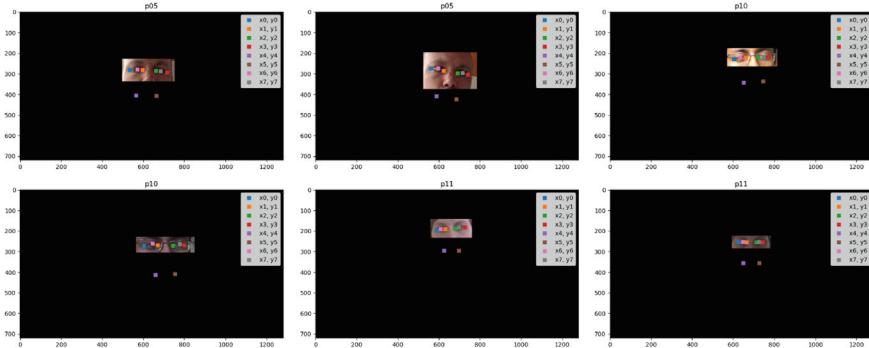
### 3 Dataset

The dataset used in this project is the MPIIGaze dataset, which is a well-known resource in the field of gaze estimation research, (see Fig. 1). It consists of pictures that show people's faces in a variety of environments, including different lighting, backdrops, and stances. Most importantly, the collection contains annotations describing where the participants' pupils are in the photos. These annotations play a crucial role in evaluating and training algorithms that try to guess where someone is looking. Subsets of the data are created, sometimes classified by users or sessions, and each subset has folders holding photos and associated annotation files. To anticipate hidden states connected to eye movements, the project's algorithm reads these annotation files, extracts relevant data including picture pathways and eye pupil locations, and then trains a Hidden Markov Model (HMM) on the data. The annotated eye locations are displayed with sample images through the use of visualizations, which offer insight into the dataset and possible uses for the built model. In the end, this dataset is a useful resource for furthering gaze estimate research, with implications for virtual reality, assistive technology, and human–computer interaction.

In this work, eye pupil coordinate extraction, (see Fig. 2), involves a computationally intensive process to accurately determine the positions of the pupils inside images. Preprocessing is done on images initially to enhance quality and for additional analysis. Next, pertinent information is extracted from the annotations in the dataset, including pupil coordinates. These coordinates are modified to ensure accurate alignment with the associated photographs. The areas that are most likely to contain the pupils are then identified using feature extraction techniques based on pixel intensities, gradients, or other visual clues. Computers then employ thresholding or edge detection techniques to precisely pinpoint the students. Validation processes are used to verify that the extracted coordinates are reliable, and refining



**Fig. 1** Defining the pupil position by visualizing the dataset



**Fig. 2** Pupil location extraction

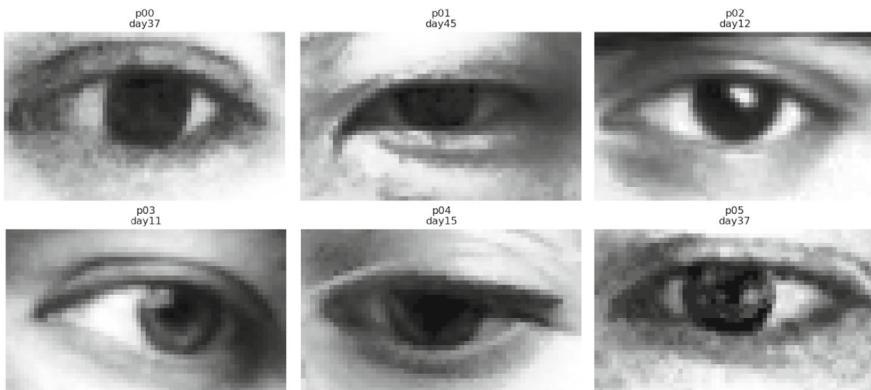
steps can be applied to increase the accuracy of the locations that are found. Consequently, this enables for further analysis such as gaze estimation since the pupil positions are saved as coordinates in the image space. Visualizations, which are often created using programs like Matplotlib, give important insights into the effectiveness of the extraction techniques employed in the study and assist validate the extraction process by presenting example pictures together with annotated pupil coordinates.

In this study, the ‘get\_eyeball’ custom function is used to extract eye pupil coordinates from the dataset ‘all\_annot\_df’. It computes the eyeball image and normalized pupil position for each data row. The extracted coordinates are then added to the data frame in the form of the newly added columns {eyeball{, {pupil\_x}, and {pupil\_y}, enabling further analysis. Using a visualization technique, (see Fig. 3), a study of the distribution of pupil coordinates and the corresponding eyeball photographs is then conducted. The next subplot grid has histograms illustrating the pupil coordinate distribution along with eyeball images for the lowest, mean, and maximum pupil coordinate values. This technique offers a comprehensive understanding of the variation in pupil positions across the dataset and their relationship with different ocular configurations, which considerably improves insights into gaze estimation research.

The normalized data from a set of “. Mat” files are handled methodically in this research work. Every file is parsed to extract the necessary vectors and envision data, and then it is formatted into a data frame for more in-depth analysis. To ensure the integrity of the final dataset, special attention is paid to managing potential mistakes detected during data processing. Important details about the images are contained in the generated DataFrame as shown in Fig. 4., tagged {all\_norm\_df}, along with related vectors and metadata like group and day identifiers.



**Fig. 3** Annotating the pupil position



**Fig. 4** Loading 213,656 images from the dataset

## 4 Methodology

### a. Problem Definition:

The project aims to develop an eye-controlled mouse application, providing an alternative input method for users to control the computer mouse cursor solely using their eye movements. This addresses the need for accessibility solutions, particularly benefiting individuals with mobility impairments or special needs who may find traditional mouse input challenging.

### b. Library Selection:

For implementing the project, several Python libraries such as OpenCV, Mediapipe, PyAutoGUI, and HMM Learn) are chosen based on their functionalities

and suitability. OpenCV is selected for its robust capabilities in image processing and computer vision tasks. Mediapipe is chosen for its specialized models, such as FaceMesh, which facilitate facial landmark detection, including eye tracking. PyAutoGUI is utilized for its ability to simulate mouse and keyboard actions, enabling cursor control. Further, HMM Learn is used for hidden Markov model (HMM) implementation in the project, providing tools for sequence prediction and analysis.

**c. Data Acquisition:**

The research work requires the acquisition of a real-time video feed from a webcam. This involves setting up the webcam hardware and configuring the software to capture video frames continuously to perform the implementation. The webcam feed serves as input data for eye tracking and cursor control based on the eye pupil position.

**d. Eye Tracking Implementation:**

The Mediapipe FaceMesh model is used to implement eye tracking. To identify facial landmarks, such as the user's eye location and eye pupil location, each frame of the camera video must be processed. In particular, the coordinates of the left eye are taken from the identified landmarks, revealing the direction of the user's look.

**e. Cursor Movement:**

Based on the model's performance, the identified eye movements are converted into equivalent cursor motions on the computer screen. This involves employing the user's gaze to determine the direction and speed of the cursor motion, as well as mapping the range of observed eye movements to the range of cursor motions.

**f. Smooth Cursor Transitions:**

To ensure a smooth user experience with an intuitive interface, a parameter is introduced to control the speed of cursor movements. This parameter helps in reducing abrupt cursor jumps and ensuring gradual transitions, enhancing the usability of the eye-controlled mouse application.

**g. Blink Detection:**

Blink detection functionality is integrated into the application to detect when the user blinks. This is achieved by monitoring changes in the distance between consecutive eye positions. Upon detecting a blink, a predefined action, such as a mouse click event, may be triggered, allowing users to perform selection actions using eye movements.

**h. Testing and Evaluation:**

The completed eye-controlled mouse application undergoes extensive testing to assess its functionality, accuracy, and usability. Testing involves verifying the performance of eye tracking, cursor control, blink detection, and overall system responsiveness. User feedback may be collected through usability testing to identify areas for improvement and validate the effectiveness of the application in meeting user needs.

## 5 Hidden Markov Model

The implementation includes an extensive set of features designed to interpret eye-tracking data and carry out image analysis operations. Fundamentally, the implementation makes use of a variety of image analysis techniques in conjunction with HMMs to analyze patterns of eye movement. The compilation of annotated eye-tracking data to train the HMM model is the first step in the implementation process. The data must be arranged in this preprocessing step so that it can be fed into the HMM for training in the proper sequence. After the data is ready, the Gaussian HMM model—which is good at capturing the underlying patterns within the eye movement sequences—is used to train the HMM using the “hmmlearn” package.

After training the model, the HMM is utilized to forecast hidden states that signify unique eye movement patterns. These anticipated states offer valuable insights into users’ navigation within their visual surroundings, thereby enriching comprehension of human behavior and cognition. Furthermore, amalgamating HMM-based analysis with image processing functions amplifies the versatility and applicability of the method across various domains.

The implementation includes an extensive set of features designed to interpret eye-tracking data and carry out image analysis operations. Fundamentally, the implementation makes use of a variety of image analysis techniques in conjunction with HMMs to analyze patterns of eye movement. The compilation of annotated eye-tracking data to train the HMM model is the first step in the implementation process. After the data is ready, the Gaussian HMM model—which is good at capturing the underlying patterns within the eye movement sequences—is used to train the HMM using the “hmmlearn” package. The use of HMM-based eye movement analysis in the study paper highlights how crucial machine learning methods are for analyzing user behavior. By deciphering the minute details of eye movement patterns, this application greatly improves convenience, medical imaging, and human-computer interaction. The effectiveness of the technique is demonstrated through experimental validation, underscoring the possibility of its practical implementation.

## 6 Cursor Control Algorithm

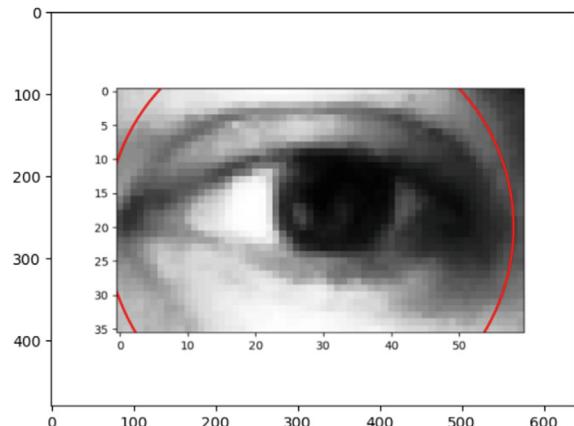
```

1      Start
2      Initialization
3          Initialize the FaceMesh model
4          Open Webcam
5          Set parameters
6              Smooth factor
7              Previous cursor position
8              Blink counter
9              Blink threshold
10     Main Loop
11         Capture frame from webcam
12         Convert frame to RGB
13         Detect faces and landmarks
14         Extract left-eye coordinates
15     Calculate Cursor Movement
16         Determine the center of the left eye
17         Calculate cursor movement, the difference in X and Y directions
18         Apply smoothing factor
19     Blink Detection
20         Measure the distance between consecutive left eye positions
21         Increment blink counter if distance < blink threshold
22     Cursor position update
23         Update cursor position based on calculated movement
24         Move cursor using PyAutoGUI
25     Cursor position update
26         Show processed frame with landmarks and cursor movement
27     Exit condition and Cleanup
28         Wait for 'ESC' key press to exit the loop
29         Release the webcam and close OpenCV windows
30     End

```

The disclosed cursor control technique tracks eye movements by using blink detection and facial landmark identification to control the cursor (see Fig. 5). First, the webcam is opened and the FaceMesh model is initialized. The main loop moves the pointer by the left eye's movements, continually records frames, and recognizes facial landmarks. To detect blink occurrences, blink detection is used, and the cursor position is adjusted correspondingly. The processed frame is shown on the screen together with the movement of the pointer and facial landmarks. When a key is pressed to break out of the loop, the algorithm ends, releasing the camera and shutting off OpenCV windows.

**Fig. 5** The picture depicts the detected eye pupil



## 7 Results

This innovative cursor control technology offers users a seamless and intuitive interaction experience with computing devices. By enabling real-time adjustment of the cursor through eye movements, users can effortlessly navigate and interact with their devices. The eye pupil positions are extracted using a Hidden Markov Model. Eye-pupil coordinates are initially obtained from images using feature extraction and localization techniques. These coordinates are then fed into an HMM, which leverages the sequential nature of eye movements to predict the next state of the pupil. By analyzing the sequence of pupil positions over time, the HMM infers underlying patterns in eye movements and predicts the most likely future position of the pupil. Subsequently, this predicted position is utilized to control the movement of the cursor on the screen.

A grayscale image, obtained from the dataset, is subjected to blob detection using the Difference of Homogeneous (DoH) method implemented in the ‘blob\_doh’ function from the ‘skimage.feature’ module. Blob detection is then performed on the binary mask with specified parameters such as ‘max\_sigma’, ‘min\_sigma’, and ‘threshold’ to identify potential blobs representing the pupil. The resulting blobs are visualized on the grayscale image using a red circle overlay (see Fig. 5), indicating the detected pupil positions. This process enables the identification and localization of the pupil within the image, facilitating further analysis of eye movement behavior. The generated visualization aids in understanding the efficacy of the blob detection method in accurately capturing the pupil’s position within the image, contributing valuable insights to the research findings.

The built eye-controlled mouse program tracks the user’s left eye’s movement in real-time using a camera feed by utilizing Python modules like OpenCV, Mediapipe, and PyAutoGUI. With the use of the program, users may operate the mouse cursor with just their eye movements by precisely translating observed eye movements into matching motions on the screen. To minimize abrupt cursor leaps and guarantee

**Fig. 6** Cursor movement position with blink counter

```
Cursor Movement (X): 78
Cursor Movement (Y): 71
Eye Distance: 0.0
Blink detected! Performing a select event.
Blink Counter: 1
```

smooth cursor transitions, a smooth factor parameter is utilized to improve the user experience. Blink detection functionality is another element of the program that lets it know when the user blinks. When the program detects a blink, it simulates a selection action by launching a mouse click event (see Fig. 6). Overall, the outcome of the implementation is a functional and intuitive eye-controlled mouse application that provides an alternative input method for users, particularly those with mobility impairments or special needs.

## 8 Conclusion

This research paper presents a novel method of eye movement-based cursor control, which represents a major advancement in the field of human–computer interaction. The system seamlessly integrates assistive technologies and advanced computer vision algorithms to let people navigate digital interfaces with ease and organicity. Its functionality depends on the accurate recognition of facial landmarks, namely the eyes. The gadget records eye movements in real-time and translates them into cursor movements using strong facial landmark verification algorithms. Enhancing the user experience and facilitating hands-free interaction by utilizing the blink detection algorithm, the system’s central component. Through monitoring changes in ocular movements and detecting blinking, the system may trigger specific actions such as mouse clicks or picks, providing users with natural control over their virtual environment. This feature is particularly helpful for those who have trouble with their motor skills since it eliminates the need for physical input devices and meets a variety of user needs and preferences.

Through the integration of machine learning techniques based on user behaviors and usage patterns, the system gains the capacity to enhance its functions through user experience modification and adjustment. Using continuous learning and improvement procedures, the system maximizes cursor control and boosts productivity, providing users with an individualized and flexible interaction experience. Thorough testing verifies the system’s functionality, demonstrating its adaptability and efficiency in a range of settings and use cases.

The research initiative establishes the groundwork for future advancements in assistive technology and accessibility. The project provides new possibilities for accessible design and digital access through a creative investigation into human–computer interaction methodologies and the application of modern technologies

such as computer vision and machine learning. Long-term research and development projects in this area might improve accessibility and give people with disabilities more power, which would eventually help to create a digital world that is more inclusive and fair.

## References

1. Sivasangari A et al (2020) Eyeball based cursor movement control. In: 2020 international conference on communication and signal processing (ICCSP). IEEE
2. Ganga V et al Eye ball movement controlled mouse and keyboard using camera. *Int J Adv Res, Ideas Innov Technol.* ISSN
3. Mangaiyarkarasi M, Geetha A (2014) Cursor control system using facial expressions for human-computer interaction. *Int J Emerg Technol Comput Sci & Electron* 8(1):30–34
4. Dhanasekar J et al (2023) System cursor control using human eyeball movement. In: 2023 3rd international conference on advance computing and innovative technologies in engineering (ICACITE). IEEE
5. Narahari D et al (2022) Eyeball movement-based cursor control for physically challenged people. *EasyChair* (7587)
6. Chandra B, Rohit M, Sriram Vignesh R (2022) Eyeball movement cursor control using OpenCV. *ECS Trans* 107.1:10005
7. Lakshmy TSS, Sivadasan A, Nitha L (2017) Detection of fraudsters in electronic auction using hidden Markov model with X-means clustering algorithm. *Int J Pure Appl Math* 114(11):157–165
8. Havarbhavi P, Tanusha G, Ashwini K (2024) Controlling of cursor movement with eyeball using hidden markov model. In: 2024 3rd international conference for innovation in technology (INOCON). IEEE
9. Sabab SA et al (2016) Eye pointer: a real time cost effective computer controlling system using eye and head movement. (c):153–159
10. Ray AK et al (2022) Cursor motion control using eye tracking and computer vision. In: International conference on advanced communication and intelligent systems. Springer Nature Switzerland, Cham
11. Narendra CSD, Gupta BNS eyeball movement based cursor control
12. Pradhan K et al (2023) Eyeball movement based cursor using machine learning. *Int Res J Mod Eng Technol Sci* 5:362–368
13. Kumar KN (2020) Analyzing cursor movements with an HMM to assess individual differences in cognition reliably and quickly. Indiana University
14. Salunkhe P, Patil AR (2015) A review on device controlled using eye movement. *Int J Emerg Trends Sci Technol* 2(1):1773–1778
15. Kim J et al (2020) A hidden Markov model for analyzing eye-tracking of moving objects: case study in a sustained attention paradigm. *Behav Res Methods* 52:1225–1243
16. Feng G (2006) Eye movements as time-series random variables: a stochastic model of eye movement control in reading. *Cogn Syst Res* 7(1):70–95
17. Aziz MF (2014) Applications of hidden Markov model and support vector machine for state estimation. University of Malaya (Malaysia)
18. Hsiao JH et al (2021) Eye movement analysis with hidden Markov models (EMHMM) with co-clustering. *Behav Res Methods* 53(6):2473–2486
19. Alhamzawi HA (2018) Control mouse cursor by head movement: development and implementation. *Appl Med Inform* 40(3/4):39–44
20. Madhuri K, Praveen Kumar L (2013) Cursor movements controlled by real time hand gestures. *Int J Sci Res* 2(2)

21. Phillips JG, Triggs TJ (2001) Characteristics of cursor trajectories controlled by the computer mouse. *Ergon* 44(5):527–536
22. Sharanyaa S, Madhumitha RP (2021) Eyeball cursor movement detection using deep learning. Proceedings of the international conference on innovative computing & communication (ICICC)
23. Ashwini K, Ponuma R, Amutha R (2021) Fine motor skills and cognitive development using virtual reality-based games in children. In: *Handbook of decision support systems for neurological disorders*. Academic Press, pp 187–201
24. Sachin Krishnan P, Rameshkumar K, Krishnakumar P (2020) Hidden Markov modelling of high-speed milling (HSM) process using acoustic emission (AE) signature for predicting tool conditions. In: *Advances in materials and manufacturing engineering: proceedings of ICAMME 2019*. Springer Singapore

# Comparative Study: Word2Vec Versus TF-IDF in Software Defect Predictions



Gaurav Sharma and Priya Singh

**Abstract** Embeddings are known for their ability to understand semantic relationships, reduce dimensionality, and identify patterns in data. These techniques are mostly used in machine learning as they are helpful and can easily be integrated into prediction models. Embedding techniques such as Word2Vec and TF-IDF are commonly used for software defect prediction tasks. While creating a defect prediction model, picking the suitable embedding method is very important. This study aims to compare the two commonly used embedding techniques, Word2Vec and TF-IDF, in the context of software defect prediction. Different sets of Java projects taken from an open-source Promise repository are used for conducting the entire analysis. A thorough evaluation was carried out by training and assessing multiple deep learning models. The efficacy of both techniques was evaluated using several evaluation metrics, including Matthews correlation coefficient, specificity, accuracy, and other significant performance indicators. Based on various metrics, the results show that TF-IDF outperforms Word2Vec, demonstrating its superiority in software defect prediction.

**Keywords** Defect prediction models · Dimensionality reduction · Embedding techniques · Performance indicators

## 1 Introduction

Software development is inherently complex, presenting various challenges. There are many companies that spend lots of time and resources fixing bugs within software. These bugs may seem small, but they can cause big problems with quality and how well the system functions. These bugs, if not taken care of, can mess up the system's workings. For the consumer, it can adversely impact the user experience and compromise system stability. The software defect prediction (SDP) technique

---

G. Sharma · P. Singh (✉)

Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi 110042, India  
e-mail: [priya.singh.academia@gmail.com](mailto:priya.singh.academia@gmail.com)

is a powerful tool to handle this situation. SDP helps to identify and prevent these faults way before they become evident in terms of consequences.

SDP is a tool that helps developers deal with software issues. It provides developers with an approach to finding defects in software. For achieving software engineering excellence, adopting such an approach is very important. SDP is done using predictive models, which focus on innovation. These innovations come from natural language processing (NLP), machine learning [1], and deep learning [2] techniques. Several hybrid models, which are the combination of more than one machine learning technique, have also been introduced to date. These models are more flexible than the traditional models and help improve predictive capability. Different embedding techniques, such as Word2Vec, GloVe, FastText, TF-IDF, and Doc2Vec, are used in SDP. These techniques play a crucial role in SDP tasks. These techniques help in converting words to numerical vectors for algorithms to capture semantic relationships present within textual data.

While these advancements have shown promise in improving SDP, there remains a lack of comprehensive comparisons between different word embedding techniques within the context of SDP. Specifically, a gap exists in understanding the relative performance of different embedding techniques employed over models for SDP tasks.

The contribution of this paper lies in the area of SDP. Specifically, it:

1. Examines how different embedding techniques have an impact on SDP tasks.
2. Compares the most commonly used embedding techniques, such as Word2Vec and TF-IDF, for SDP tasks based on several evaluation metrics.
3. Evaluates the performance of various deep learning models' used for SDP.
4. Analyzes how the combination of embedding techniques and deep learning models sync together to enhance SDP.
5. Provides insights for selecting the best suitable embedding techniques for SDP tasks.

The paper is organized in a structured and comprehensive manner. In the subsequent sections, we delve into the methodology employed, providing a systematic explanation of our approach. This includes all the steps performed during the experiment sequentially, and every step has been explained in a detailed manner. Then the experimental results are presented with a technical discussion of our findings and finally concluded with reflections on future research directions.

## 2 Related Work

Word embeddings work by representing text in n-dimensional space. They are essential for solving NLP-related problems. One such problem in identifying Swahili Smishing communications directed at mobile money customers is emphasized by Iddi et al. [3]. These techniques allow for efficient classification by capturing semantic links in text, which is crucial for differentiating genuine messages from smishing ones. Similarly, to establish a unified feature space for text and image modalities,

Zongwei [4] introduced a multi-modal approach by integrating TF-IDF features with LSTM networks for capturing sequential information. The incorporation of TF-IDF helps in refining the image modalities.

Emotion processing is becoming an important research area in fields such as data analysis and NLP. For analyzing emotion, it is important to capture the presence of specific words along with their relationships with other words. Sabery [5] proposed a hybrid model for emotion analysis by combining the Deep Belief Network with TF-IDF and Glove. The embeddings helped outperform the baseline models in several metrics. In a similar context of emotion analysis, Canales [6] achieved efficient data annotation through Word2Vec embeddings, enhancing the categorization process of different emotions.

In the context of SDP, the role of defect prediction models becomes equally important as the embedding techniques. For a longer period of time, conventional machine learning approaches have been used in creating prediction models. The issue with these approaches lies in capturing the semantic relationship among textual data. In comparison to these approaches, neural networks perform better with image and textual data. Using a deep learning approach, Miholca [7] significantly improved defect prediction, outperforming conventional methods in the Calcite program. The significance of SDP is also highlighted by Nevendra [8] in concerns regarding software complexity. The research shows notable performance gains by comparing deep learning techniques across open-source projects. This change in strategy creates new opportunities for improving defect prediction models.

Using hybrid features [9] is advantageous, making defect prediction models more flexible. Wang cleverly combined the AST and Control Flow Graph (CFG) via the Graph Isomorphism Network to push SDP with H-GIN as evidence. With respect to the PyTraceBus dataset, H-GIN demonstrated better prediction accuracy than earlier approaches. Similarly, graph neural networks (GNN) and transformers were used to create a novel model that Tang and He presented [10]. Their technique, which included absolute and relative locations in the AST, addressed the local learning constraints of GNN and showed superior F-measure and improved detection of faulty features on the PROMISE dataset.

A dataset of more than 400 thousand articles from design pattern books was used by D. Liu et al. for DPWord2Vec [11], a technique to concurrently embed design patterns and natural language words into vectors. According to evaluation, DPWord2Vec performs 24.2–120.9% better than baseline algorithms when assessing word and design pattern similarity. Additionally, DPWord2Vec enhances design pattern tasks by 6.5–70.7%, including tag suggestion and selection. A similar approach to learning from datasets was used in a technique for proposing Web services for superior Mashup applications put forth by Cao et al. [12]. Their method uses Word2Vec for semantic representations from service descriptions and creating a service relationship network, combining bilinear graph attention representation with xDeepFM quality prediction. The findings on the ProgrammableWeb dataset demonstrate better performance in terms of accuracy and recall compared to other approaches.

### 3 Methodology

The methodology of this research encompasses several key stages aimed at effectively analyzing Java code for bug prediction. The stages are described in a detailed manner in the coming subsections as follows.

#### 3.1 Corpus Generation Using AST

The Python library javalang is used to represent the Java code in a tree-like structure that is the AST of the code. The Java code is taken from different Java projects described in Sect. 4.1. The Javalang library may be obtained from <https://github.com/c2nes/javalang>. There are two components in it: a lexer and a parser made specifically for Java [13]. Within the AST, each node corresponds to a specific construct such as MethodDeclaration, IfStatement, or VariableAccess, pinpointing occurrences within the source code. As a result, the AST facilitates the generation of a comprehensive corpus for each Java project. This corpus is used for fine-tuning the pre-trained TF-IDF and Word2Vec models.

#### 3.2 Generation of Sequence Tokens

The categories of AST nodes selected as tokens are control flow nodes, class declarations, and method invocations, which are also depicted in Table 1. A new sequence token file is created for every version of the Java project (for example, Ant 1.5), and when any of the selected tokens in the table is detected within the corpus generated by the AST, that token is appended to the sequence token file. This procedure iterates for every version of the Java project, thereby composing the sequence tokens. These tokens are subsequently utilized as input for the models to generate embeddings.

**Table 1** AST selected nodes

MethodInvocation	SuperMethodInvocation	PackageDeclaration	InterfaceDeclaration
ClassDeclaration	MethodDeclaration	ConstructorDeclaration	VariableDeclarator
FormalParameter	BasicType	CatchClauseParameter	MemberReference
SuperMemberReference	ReferenceType	IfStatement	WhileStatement
DoStatement	ForStatement	AssertStatement	BreakStatement
ContinueStatement	ReturnStatement	ThrowStatement	SynchronizedStatement
TryStatement	SwitchStatement	BlockStatement	StatementExpression
TryResource	CatchClause	CatchClauseParameter	SwitchStatementCase

### 3.3 Fine Tuning of Pre-Trained Model

Transfer learning is employed by importing Word2Vec and TF-IDF models from Gensim and Scikit-Learn libraries, respectively. The models are trained on the corpus generated by AST for each Java project. These trained models are fed with tokens to generate the embeddings. TF-IDF and Word2Vec models are widely used statistical methods in NLP. TF-IDF measures how important a term is within a document relative to a collection of documents. On the other hand, Word2Vec represents words as dense vectors in a continuous space using shallow neural networks.

TF-IDF Vectorizer, imported from scikit-learn, is trained on the corpus. The vectorizer is fitted to the data using ‘fit-transform()’, analyzing text, constructing vocabulary, and calculating TF-IDF scores. The resulting sparse matrix represents documents, words, and TF-IDF scores, forming the trained TF-IDF model. Similarly, the Word2Vec model is imported from the Gensim library to train on a corpus generated by AST. The model is initialized and trained with specific parameters, such as a vector size of 100 words, a window size of 5, a minimum count of 5, and a number of epochs for training the model of 10.

### 3.4 Generation of Embeddings

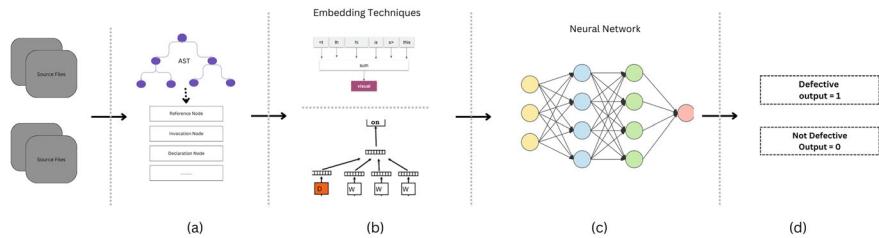
After subjecting the pre-trained Word2Vec and TF-IDF models to fine-tuning with the corpus generated by AST, the sequence tokens extracted from each version of the project are then inputted into the trained models. By doing so, the models are able to produce embeddings that reflect the underlying contextual information embedded within the code.

### 3.5 Workflow and the Deep Learning Models Used

The whole process, as depicted in Fig. 1, goes by training a deep learning model using the vector representations obtained from TF-IDF or Word2Vec. This model is geared towards executing a specific task of defect prediction. The training process starts by inputting the embeddings into the model and iteratively refining model parameters to enhance performance.

The models used in this research are as follows:

1. LSTM: It enables early detection of defects, reducing maintenance costs and improving quality assurance [14].
2. GRU: It is akin to LSTM but unlike LSTM, GRU architectures offer a streamlined approach to defect detection for enabling early defect detection.
3. ANN: Since these models can discover intricate patterns and correlations in the data, they are still a good option for SDP.



**Fig. 1** The overall process **a** Parsing the java code using AST. **b** Creating vectors using embedding techniques. **c** Training of deep learning model **d** Performing defect prediction

4. CNN: These are seamlessly integrated with other neural network architectures for enhanced performance across various tasks.

### 3.6 Comparison of Techniques

The output of the trained deep learning model assigns “1” for bugs detected and “0” for bug-free software. After getting the final output, a comparative analysis is conducted to evaluate the embeddings based on the evaluation metrics.

This step assesses which embedding technique contributes more effectively to the model’s performance. The important performance indicators used in this research are described in Sect. 4.2.

## 4 Experimental Setup

This section outlines the dataset utilized, the baseline deep learning model selected for comprehensive analysis, the hyperparameter setting, as well as the evaluation measures included in the experiment.

### 4.1 Dataset Used

The dataset used in this research is a set of 10 open-source Java projects that are taken from the PROMISE repository. The list of Java projects is given with their descriptions in Table 2.

**Table 2** Description of Projects along with Versions

Projects	Versions	Description
Ant	1.5, 1.6, 1.7	Java tool for managing processes, targets, and dependencies
Camel	1.2, 1.4, 1.6	Open-source Java framework simplifying integration with EIPs, diverse transports, APIs
Ivy	1.4, 2.0	Ivy, a sub-project of Apache Ant, resolves project dependencies using external XML files and downloads resources from repositories.
jEdit	4.0, 4.1, 4.2, 4.3	jEdit offers native syntax highlighting for more than 200 file formats, extendable via XML. Supports UTF-8 and various encodings with robust folding and wrapping features
Log4j	1.0, 1.1, 1.2	log4j is integral to the Apache Logging Services Project, offering dependable, open-source logging utilities for diverse application needs
lucene	2.0, 2.2, 2.4	Lucene is a Java-based, high-performance search engine library ideal for applications needing structured or full-text search, facetting, and more
poi	2.0, 2.5, 3.0	POI, an open-source Java library, facilitates creation and manipulation of Microsoft Office file formats, enabling operations such as creation, modification, and reading
synapse	1.1, 1.2	Synapse is analytics service merging data warehousing and Big Data analytics, offering flexible querying options with scalable resources
xalan	2.4, 2.5, 2.6, 2.7	Xalan-Java functions as an XSLT processor, converting XML documents into various formats such as HTML, text, or other XML document types
xerces	1.2, 1.3	It incorporates the Xerces Native Interface (XNI), offering a highly modular and programmable framework for building parser components and configurations

## 4.2 Evaluation Measure

A variety of evaluation metrics were employed as assessment measures to provide a full examination of the model's performance across varied criteria. In the following equations, there are several key terms, which are mentioned below:

1. True Positive (TP): Instances correctly classified as positive by the model.
2. False Positive (FP): Instances incorrectly classified as positive by the model.
3. True Negative(TN): Instances correctly classified as negative by the model.
4. False Negative(FN): Instances incorrectly classified as negative by the model.

Precision: It is the ratio of correctly predicted positive outcomes to all of the model's predicted positive outcomes.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

Recall: also known as sensitivity or true positive rate, measures the proportion of the actual positive values that were identified accurately by the model.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

F1 Score: The harmonic mean of recall and accuracy is F1 Score.

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

Accuracy: It is the percentage of properly categorised examples among all the instances.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \quad (4)$$

MCC: a measure of the effectiveness of binary classification models, especially in the case of unbalanced datasets.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (5)$$

FNR: False Negative Rate is a measure that represents the proportion of actual positive instances that were incorrectly classified as negative by a model.

$$\text{FNR} = \frac{\text{FN}}{\text{TP} + \text{FN}} \quad (6)$$

FPR: False Positive Rate indicates the percentage of real positive examples that a model misclassified as negative.

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \quad (7)$$

TNR: True Negative Rate shows the percentage of real negative cases that a model accurately identified as negative.

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (8)$$

### 4.3 Hyperparameter Settings

Training spanned 200 epochs to ensure comprehensive data learning. For ANN architectures, a sigmoid activation function was utilized throughout the layers, while rectified linear unit (ReLU) activation was applied in dense layers of RNNs (LSTM and GRU) with sigmoid activation in the output layer. Binary cross-entropy served as the loss function across all models, optimized by the Adam optimizer. A batch size of 32 was chosen for computational efficiency and model stability. Two dense layers with 64 and 32 neurons, respectively, were employed to capture intricate data patterns. For CNN architecture, 1D convolutional layers were leveraged to capture spatial dependencies in sequential software data, thereby enhancing overall model performance.

## 5 Result and Discussion

In this section, the performance of TF-IDF and Word2Vec across the four discussed deep learning models is presented. Tables 3, 4, 5, 6 contain the projects on which the models are trained, along with the mean values of accuracy (Acc.), precision (Prec.), F1-score (F1), MCC, and TNR, calculated from various versions of the same projects. For instance, the Lucene project had versions 2.0, 2.2, and 2.4. Training on version 2.0 and testing on version 2.2 yielded an accuracy of 0.63, while training on version 2.2 and testing on version 2.4 resulted in an accuracy of 0.61. The mean accuracy, calculated as 0.62, is included in the table. Detailed metrics are available at [github.com/GauravSharma171691](https://github.com/GauravSharma171691). Figure 2a and b show mean FPR and FNR values for both TF-IDF (T) and Word2Vec (W).

**Table 3** Comparison of CNN based models using Word2Vec and TF-IDF embeddings

Projects	Word2Vec					TF-IDF				
	Prec	F1	Acc.	MCC	TNR	Prec	F1	Acc.	MCC	TNR
Ant	0.237	0.380	0.241	0.062	0.014	0.252	0.401	0.281	0.086	0.050
Camel	0.188	0.315	0.195	0.070	0.007	0.285	0.336	0.715	0.193	0.731
Ivy	0.263	0.413	0.275	0.014	0.027	0.170	0.270	0.102	0.012	0.017
jEdit	0.131	0.235	0.139	0.014	0.020	0.199	0.327	0.418	0.207	0.471
log4j	0.650	0.735	0.663	0.082	0.029	0.658	0.729	0.676	0.074	0.038
lucene	0.621	0.764	0.624	0.107	0.009	0.615	0.759	0.616	0.068	0.022
poi	0.469	0.594	0.475	0.026	0.012	0.478	0.613	0.481	0.051	0.022
synapse	0.274	0.429	0.280	0.016	0.014	0.258	0.408	0.257	0.111	0.000
xalan	0.649	0.761	0.648	0.026	0.002	0.649	0.761	0.649	0.007	0.003
xerces	0.443	0.638	0.442	0.044	0.011	0.217	0.330	0.215	0.061	0.020

**Table 4** Comparison of ANN models using Word2Vec and TF-IDF embeddings

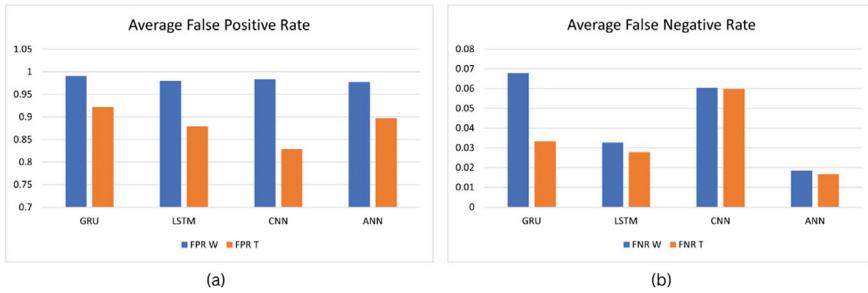
Projects	Word2Vec					TF-IDF				
	Prec	F1	Acc.	MCC	TNR	Prec	F1	Acc.	MCC	TNR
Ant	0.256	0.408	0.253	0.086	0.018	0.257	0.409	0.253	0.086	0.019
Camel	0.189	0.313	0.197	0.058	0.012	0.171	0.283	0.168	0.060	0.007
Ivy	0.085	0.156	0.117	0.024	0.035	0.099	0.166	0.123	0.029	0.023
jEdit	0.132	0.231	0.310	0.028	0.034	0.132	0.237	0.314	0.075	0.024
log4j	0.651	0.732	0.660	0.126	0.036	0.651	0.732	0.660	0.126	0.036
lucene	0.614	0.759	0.614	0.021	0.006	0.614	0.759	0.614	0.021	0.006
poi	0.468	0.612	0.475	0.041	0.012	0.468	0.612	0.475	0.041	0.012
synapse	0.277	0.432	0.290	0.027	0.028	0.277	0.432	0.290	0.027	0.028
xalan	0.649	0.761	0.649	0.007	0.003	0.649	0.761	0.649	0.007	0.003
xerces	0.166	0.284	0.165	0.132	0.000	0.166	0.284	0.165	0.132	0.000

**Table 5** Comparison of GRU models using Word2Vec and TF-IDF embeddings

Projects	Word2Vec					TF-IDF				
	Prec	F1	Acc.	MCC	TNR	Prec	F1	Acc.	MCC	TNR
Ant	0.241	0.392	0.237	0.083	0.015	0.276	0.426	0.276	0.126	0.105
Camel	0.186	0.309	0.188	0.043	0.001	0.184	0.306	0.184	0.038	0.002
Ivy	0.096	0.160	0.126	0.026	0.050	0.083	0.153	0.138	0.046	0.035
jEdit	0.130	0.230	0.181	0.057	0.021	0.157	0.263	0.150	0.070	0.128
log4j	0.644	0.758	0.641	0.031	0.009	0.651	0.779	0.649	0.083	0.025
lucene	0.612	0.757	0.609	0.051	0.001	0.613	0.757	0.611	0.048	0.000
poi	0.469	0.613	0.470	0.175	0.012	0.507	0.637	0.509	0.198	0.163
synapse	0.256	0.407	0.259	0.079	0.009	0.269	0.423	0.289	0.029	0.040
xalan	0.482	0.645	0.482	0.019	0.002	0.482	0.648	0.481	0.014	0.001
xerces	0.168	0.287	0.176	0.026	0.013	0.153	0.263	0.184	0.054	0.045

**Table 6** Comparison of LSTM models using Word2Vec and TF-IDF embeddings

Projects	Word2Vec					TF-IDF				
	Prec	F1	Acc.	MCC	TNR	Prec	F1	Acc.	MCC	TNR
Ant	0.240	0.387	0.240	0.090	0.007	0.280	0.435	0.387	0.197	0.199
Camel	0.175	0.282	0.168	0.018	0.005	0.193	0.319	0.247	0.048	0.089
Ivy	0.088	0.158	0.142	0.025	0.036	0.084	0.154	0.101	0.009	0.021
Jedit	0.130	0.222	0.133	0.040	0.005	0.183	0.286	0.432	0.237	0.360
log4j	0.653	0.780	0.642	0.097	0.051	0.645	0.736	0.641	0.078	0.009
lucene	0.617	0.759	0.614	0.076	0.003	0.623	0.758	0.621	0.086	0.068
poi	0.470	0.636	0.475	0.028	0.012	0.494	0.607	0.521	0.149	0.143
synapse	0.270	0.427	0.275	0.051	0.007	0.279	0.429	0.352	0.106	0.147
xalan	0.647	0.759	0.647	0.021	0.002	0.651	0.763	0.651	0.029	0.004
xerces	0.186	0.315	0.170	0.013	0.014	0.168	0.287	0.176	0.026	0.013



**Fig. 2** Average values of **a** False positive rates **b** False negative rates across all the projects

**Table 7** Average performance of deep learning models using Word2Vec and TF-IDF

Embedding	Model	Precision	F1Score	Accuracy	MCC
Word2Vec	GRU	0.3809	0.4976	0.3860	0.0668
	LSTM	0.3795	0.4927	0.3836	0.0506
	CNN	0.3920	0.5121	0.3978	0.0610
	ANN	0.3826	0.4989	0.3893	0.0551
TF-IDF	GRU	0.3947	0.5099	0.4259	0.0879
	LSTM	0.3998	0.5167	0.4618	0.1100
	CNN	0.4047	0.5166	0.5087	0.0974
	ANN	0.4004	0.5153	0.4584	0.0848

With the help of tables, models with TF-IDF embeddings exhibit superior performance compared to those leveraging Word2Vec embeddings. Models with TF-IDF embeddings have higher precision, F1-score, accuracy, MCC, and TNR. Higher values of Precision, F1 Score, Accuracy, and MCC show that the model is better and effective at classification tasks. Table 7 illustrates the average performance of deep learning models when employed with Word2Vec and TF-IDF embeddings, respectively. It is evident that the models exhibit superior performance when trained with TF-IDF embeddings. For example, CNN model when fed with TF-IDF embeddings has higher precision, F1 Score, Accuracy and MCC than when it is fed with Word2Vec embeddings. This represents that TF-IDF embeddings help the model to perform better at classification task.

The performance of the models trained in this study heavily depends on various aspects of the data it is trained on.

1. Too small dataset can lead to poor performance of model on unseen data.
2. Too large dataset may create models with many parameters affecting the deployment of model.

In this research, within-project defect prediction (WPDP) is used. In WPDP, the models are trained on a version of the project and tested on another subsequent version of the same project. This situation's performance can be improved by using

cross-project defect prediction (CPDP) [15], where the model can be trained on one project and tested on another project.

## 6 Conclusion and Future Scope

The performance of two different word embedding methods, TF-IDF and Word2Vec, is assessed in the context of SDP using different deep learning methods. The main aim was to assess the effectiveness of TF-IDF and Word2Vec across various performance indicators. The findings offer strong evidence that TF-IDF embeddings are superior to Word2Vec embeddings in categorization on a constant basis, as TF-IDF regularly outperforms Word2Vec in assessment measures. These measures signify the models with TF-IDF embeddings have the ability to correctly classify instances, have fewer miss-classifications, capture overall correlation, and accurately identify true negative cases, respectively, thereby indicating superior predictive capability.

This research opens up several promising directions for future work. Firstly, there is potential to explore more advanced and specialized embedding techniques such as FastText, Doc2Vec, BERT, Code-BERT, ROBERTA, ELMO, and XLNET. These techniques have shown promising results in capturing semantic relationships with better accuracy. The dataset used in this research is taken from the PROMISE dataset, which can also be replaced with NASA's dataset for software engineering, which has the capability of producing a larger corpus. CPDP can also be used in place of WPDP so as to increase the efficacy of deep learning models. Furthermore, hybrid models can be used for better performance. Additionally, the results may be more broadly applicable if these enhanced models and embeddings are used in a wider variety of languages and domains. With this generalization, SDP would become more resilient and adaptable. It is important to constantly benchmark against new models and embedding techniques. It ensures that advancements stay at the forefront of the industry.

## References

1. Sharma T, Jatain A, Bhaskar S, Pabreja K (2023) Ensemble machine learning paradigms in software defect prediction. *Proc Comput Sci* 218:199–209. <https://doi.org/10.1016/j.procs.2023.01.002>
2. Malhotra R, Singh P (2023) Recent advances in deep learning models: a systematic literature review. *Multimed Tools Appl* 82:44977–45060. <https://doi.org/10.1007/s11042-023-15295-z>
3. Mambina IS, Ndibwile JD, Michael KF (2022) Classifying Swahili smishing attacks for mobile money users: a machine-learning approach. In: *IEEE Access*, pp 83061–83074. <https://doi.org/10.1109/ACCESS.2022.3196464>
4. Es-Sabery F, Es-Sabery I, Hair A, Sainz-De-Abajo B, Garcia-Zapirain B (2022) Emotion processing by applying a fuzzy-based Vader Lexicon and a Parallel deep belief network over massive data. *IEEE Access* 10:87870–87899

5. Xie Z, Liu L, Wu Y, Li L, Zhong L (2021) Learning TF-IDF enhanced joint embedding for recipe-image cross-modal retrieval service. IEEE. Publisher, IEEE, pp 3304–3316
6. Canales L, Strapparava C, Boldrini E, Martínez-Barco P (2020) Intensional learning to efficiently build up automatically annotated emotion corpora. *IEEE Trans Affect Comput* 11(2):335–347. <https://doi.org/10.1109/TAFFC.2017.2764470>
7. Miholca D-L, Tomescu V-I, Czibula G (2022) An in-depth analysis of the software features' impact on the performance of deep learning-based software defect predictors. *IEEE Access* 10:64801–64818
8. Nevendra M, Singh P (2022) A survey of software defect prediction based on deep learning. *Archiv Comput Methods Engin* 29:5723–5748. <https://doi.org/10.1007/s11831-022-09787-8>
9. Wang X, Lu L, Wang B, Shang Y, Yang H (2023) SDP via GIN with hybrid graphical features. In: IEEE 22nd international conference on software quality, reliability, and security companion (QRS-C)
10. Tang F, He P (2023) SDP using multi-scale structural information. In: ICCAI '23: proceedings of the 2023 9th international conference on computing and artificial intelligence, March 2023, pp 548–556
11. Liu D, Jiang H, Li X, Ren Z, Qiao L, Ding Z (2022) DPWord2Vec: better representation of design patterns in semantics. *IEEE Trans Softw Engin* 48(4):1228–1248
12. Cao B, Zhang L, Peng M, Qing Y, Kang G, Liu J (2023) Web service recommendation via combining bilinear graph representation and xDeepFM quality prediction. *IEEE* 20(2):1078–1092
13. Fan G, Diao X, Yu H, Yang K, Chen L (2019) Software defect prediction via attention-based recurrent neural network. *Sci Program.* <https://doi.org/10.1155/2019/6230953>
14. Liang H, Yu Y, Jiang L, Xie Z (2019) SEML: a semantic LSTM model for software defect prediction. *IEEE Access* 7:83812–83824. <https://doi.org/10.1109/ACCESS.2019.2925313>
15. Bala YZ, Samat PA, Sharif KY, Manshor N (2022) Improving cross-project SDP method through transformation and feature selection approach. *IEEE Access* 11:2318–2326. IEEE

# An Improved Deep Learning Framework Based on Multi-Scale Convolutional Architecture for Road Crack Detection



Idris Ya'u Idris , Badamasi Imam Ya'u , Usman Ali , and Yonis Gulzar

**Abstract** In addition to serving as a means of transportation, a highway connects the local economy. But as the pavement ages, several flaws including cracks, potholes, and deformation gradually show up on the surfaces of the road. Crack evaluation is typically done by hand using field surveys performed by humans. Nevertheless, these labor-intensive, time-consuming, manual survey methods are not reproducible or repeatable and place surveyors in dangerous situations. This proposes a novel deep-learning framework for road crack detection based on multi-scale convolutional architecture. To address the challenges of object occlusion, soft non-maximal suppression (NMS) is applied across crack proposals at various feature scales. The experimental results show that the introduced model attains an accuracy of 0.9989 and better detection performance across four distinct object proposal building block (OPBB) architectures. Thus, our model outperforms the benchmark models in every instance of OPBBs. In this way, the problem of large object scale variation in road crack images with different object occlusion challenges has been tackled, and the positive images are divided into a set of semantically significant regions, such as the road surface and cracks.

**Keywords** Road cracks · Deep learning · Convolutional architecture

---

I. Y. Idris

Department of Computer Science, School of Science and Technology, Federal Polytechnic Bauchi, Bauchi 743001, Nigeria

B. I. Ya'u

Department of Mathematical Sciences, Faculty of Science, Abubakar Tafawa Balewa University, Bauchi 740272, Nigeria

U. Ali

Department of Computer Science, School of Science, Federal College of Education (Technical), Gombe 760253, Nigeria

Y. Gulzar

Department of Management Information Systems, College of Business Administration, King Faisal University, Al Ahsa 31982, Saudi Arabia  
e-mail: [ygulzar@kfu.edu.sa](mailto:ygulzar@kfu.edu.sa)

## 1 Introduction

The state of roads, a vital element of city infrastructure, has a direct effect on the daily experiences of residents [1, 2]. Minor road diseases may eventually worsen and cause major damage if the appropriate departments ignore this issue. This could result in traffic jams, an early end to the life of the road, and even significant risks to public health and safety [3, 4]. Numerous academics have conducted pertinent research and proposed a range of monitoring techniques in response to the challenges associated with road management and maintenance. These techniques include the use of high-resolution satellite images, radar, and laser scanning to assess road damage [3, 5].

Many types of structures, including pavements, buildings, and bridges, exhibit cracking. According to [6], cracking can hasten the process of deterioration, so its frequency and severity are crucial markers that maintenance is necessary. Thus, one of the most important tasks in ensuring public safety is crack evaluation. Traditionally, human field surveys are used to manually evaluate cracks [7, 8]. Through the management process, crack diagnosis can be used to establish the best course of action in terms of timing and intensity of treatment [9, 10].

Recent research into characterizing pavement surface distress and detecting cracks has seen a surge of interest [11, 12]. Traditionally, methods for crack detection involve assigning gradient features to each pixel in an image, which are then processed by binary classifiers to identify crack presence. Different approaches have been explored, such as using the Gabor filter [13–15] and local binary patterns (LBPs) [16]. CrackTree, an automated system employing a tree structure for crack detection, was introduced in [17, 18]. Additionally, integrated systems for comprehensive crack detection and characterization have been proposed [19, 20]. Moreover, various image processing algorithms have been developed specifically for detecting and characterizing crack distress on road pavements [21, 22]. According to [7], deep learning, particularly in comparison to traditional machine learning, has emerged as a dominant technology for advancing crack detection algorithms. Deep learning's superiority in object detection and semantic segmentation has propelled its prominence in crack detection research [23, 24]. It represents an extension and enhancement of traditional machine learning approaches [14, 25–30].

Deep learning-based object detection technology has been progressively improved in road traffic in recent years. This includes the identification of traffic congestion, the recognition of vehicle and license plate numbers, and the detection of pavement damage [3]. Among them, deep learning was initially used in [31] for pavement damage detection.

Detecting road cracks is essential for maintaining road safety and preventing both accidents and damage to infrastructure [23, 24]. Conventional techniques for identifying cracks in roads usually rely on human inspectors. This approach is inefficient, requiring significant time and manpower, and is vulnerable to errors due to human limitations [16]. Additionally, these methods find it challenging to manage extensive road networks and the large amounts of data that must be analyzed. Consequently,

there is an urgent need for automated and efficient systems for road crack detection [6, 7].

To address these issues, researchers have developed a sophisticated artificial intelligence system. This system uses a deep learning model with a multi-layered convolutional structure capable of analyzing road surfaces at various scales, offering a potential breakthrough in automating road crack identification. This framework takes advantage of deep learning, a branch of artificial intelligence known for its exceptional performance in image recognition and segmentation tasks. By applying deep learning techniques, the proposed framework seeks to transform road crack detection by providing improved accuracy, efficiency, and scalability.

Our research advances the field by introducing three enhancements to convolutional neural network (CNN) methods for detecting road cracks visually. We suggest using deconvolution and merging CNN feature maps to incorporate more complex features and contextual information. This approach aims to improve crack detection at smaller scales within the feature maps, tackling the problem of wide variations in crack sizes. The utilization of a multi-scale convolutional architecture is particularly advantageous in this context. This architecture enables the model to capture intricate features at different scales, ranging from small cracks to larger structural patterns, thereby enhancing the overall robustness and effectiveness of the detection system. By incorporating multi-scale information, the framework can accurately delineate cracks amidst various road surface textures, lighting conditions, and environmental factors.

## 2 Related Work

The study [32] aims to detect cracks in concrete roads under various conditions using a deep learning-based object detection method. Utilizing a pre-trained Faster R-CNN, the method analyzes 323 high-resolution images. Results show consistent crack detection across different weather conditions, but accuracy drops significantly during low-light times, particularly between 7:00 and 8:00 pm.

Research by [33] presents an innovative visual analysis technique for precisely identifying both sealed and unsealed cracks in asphalt roads, while minimizing incorrect identifications. This approach integrates image analysis for extracting key features with a pattern recognition system using a support vector machine, which is fine-tuned by a Salp Swarm Algorithm. The method's effectiveness was demonstrated through experiments, yielding 91.33% accuracy in detecting cracks and 92.83% accuracy for sealed crack identification.

The research conducted in [24] examines the performance of four existing convolutional neural network (CNN) models, previously trained for other tasks, in identifying and categorizing cracks on highways. Additionally, the researcher designs a new CNN model aimed at improving the accuracy of crack detection and classification. Using 4,663 classified images, the models were tested with MATLAB, achieving

top accuracy with GoogleNet (89.08%). The newly developed CNN model, optimized with Adam's algorithm at a 0.001 learning rate, achieved 97.62% accuracy, surpassing all pre-trained models.

The study [34] proposes a lightweight YOLO-based detection algorithm to improve road pit defect detection. The model's performance is improved by integrating two key components: a Bidirectional Feature Pyramid Network (BiFPN), which enhances the extraction of relevant features, and Varifocal Loss, which helps balance the representation of different types of samples. These additions result in a more accurate model overall. Tested on the PCD1 dataset, the BV-YOLOv5S model shows mAP@0.5 improvements of 4.1, 3, and 0.9% over YOLOv3-tiny, YOLOv5S, and B-YOLOv5S models, respectively, proving its superior performance and reliability for real-time road defect detection.

Another study [35] addresses the limitations of traditional and machine learning-based pavement crack detection methods by using deep learning, specifically the YOLOv5 series models. The dataset comprises 3,001 high-resolution asphalt crack images categorized by severity. Tests reveal that among the models evaluated, YOLOv5l performs best in terms of accuracy, correctly identifying cracks 88.1% of the time. On the other hand, YOLOv5s stands out for its speed, processing each image in just 11.1 ms, making it the quickest option.

Similarly in study [36], an automated framework for crack and pothole segmentation using stereo vision and deep learning was developed. Utilizing a multi-view stereo imaging system, datasets include color, depth, and color-depth overlapped images. The researchers propose an adapted U-net design that uses depthwise separable convolution to lower processing requirements. When tested on asphalt road surfaces, this model demonstrates precision at the millimeter scale and performs exceptionally well in segmentation tasks. This high level of accuracy allows for precise measurements of pothole volumes.

In another research [37], it highlights the importance of improving road quality through automated pavement defect detection, significantly reducing monitoring time. It reviews global approaches and existing datasets for road defect detection and segmentation. The study uses deep learning with synthetic training data to segment cracks in driver-view images. The novelty lies in the synthetic dataset creation, addressing challenges like varied pixel intensity, complex crack topology, and diverse illumination conditions.

In their study, Hsieh and Tsai [7] examine the progress of machine learning (ML) and deep learning (DL) in automating crack detection. They analyzed 68 ML-based approaches to identify patterns and tested 8 models using 3D road surface images in different scenarios. Their results show that fully convolutional network (FCN) models with more complex backbone structures and U-Net models with skip connections performed better. Both these approaches scored above 90 on the enhanced Hausdorff distance metric for most categories, with the exception of the “Other Distress” group. Addressing false positives remains a crucial challenge for further improvement.

Praticò and colleagues [38] introduce a machine learning approach that uses supervision to identify and categorize the condition of road surfaces. Their method

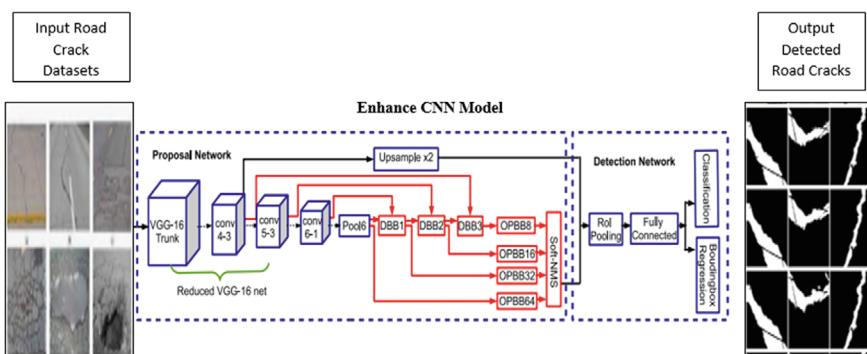
analyzes vibration and sound patterns produced by the pavement to assess its structural integrity. They utilize acoustic sensors positioned roadside to collect these signatures and employ various ML classifiers (MLP, CNN, RFC, SVC). Results indicate high accuracy in associating vibro-acoustic signatures with different types of road pavement cracks (MLP = 91.8%, CNN = 95.6%, RFC = 91.0%, SVC = 99.1%). These findings support future applications in real-world settings like monitoring roads and bridges with wireless sensor networks.

Feng et al. [39] introduce a method combining SSD and U-Net models for accurate pavement crack identification, classification, segmentation, and geometric parameter calculation. The approach enhances crack classification accuracy (transverse: 86.8%, longitudinal: 87.6%, alligator: 85.5%) by integrating crack detection and segmentation networks. Improved feature extraction and optimized hyperparameters contribute to better results, providing precise category, positioning, and geometric details crucial for evaluating pavement conditions.

### 3 Methodology

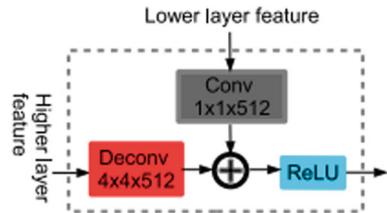
To improve crack detection at low feature map scales, deconvolution, and fusion of CNN identified in the existing study [40], feature maps are proposed as a solution to the large object scale variation problem. These techniques add context and deeper features. Moreover, to tackle the issue of object occlusion [40], we apply soft NMS across crack proposals at different feature scales. We analyze the aspect ratio distributions of different road types, leveraging their distinctive characteristics to properly configure anchor boxes for enhanced crack alignment and localization.

Experiments conducted over the road crack dataset are used to evaluate the proposed CNN enhancements with different image input sizes. These methods are believed to contribute to creating more discriminative frameworks that perform better (see Fig. 1).



**Fig. 1** Architecture of the proposed model

**Fig. 2** Feature fusion method for deconvolution building block (DBB)



The adapted CNN model processes input images with dimensions  $H \times W \times D$ , where  $H$  and  $W$  denote the image height and width in pixels, and  $D$  represents the number of color channels. Figure 1 illustrates the core structure of this modified CNN model, which is based on MS-CNN. It identifies candidate objects across multiple feature output layers of different scales. The enhancements, highlighted in red boxes in Fig. 1, distinguish it from the standard MS-CNN and are applicable to other CNN architectures like Faster-RCNN and SSD.

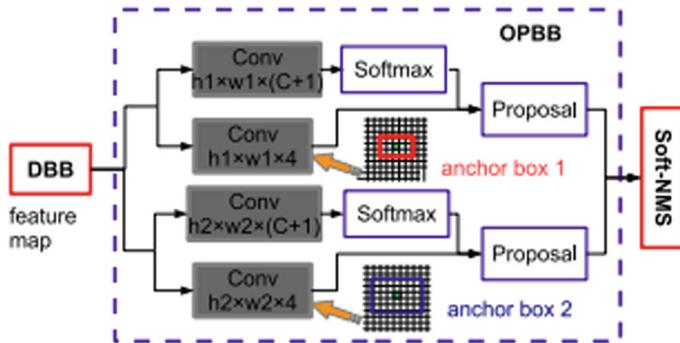
The suggested system functions as a two-part object detection model. It consists of two main components: one network that suggests potential objects of interest, and another that confirms and precisely identifies these objects. The proposed model extends from the modified VGG-16 model, incorporating additional convolution, pooling, deconvolution, and object proposal layers integrated into the original VGG-16 structure.

The system improves its initial feature analysis using specialized components called deconvolution building blocks (DBB), marked as “DB1,” “DB2,” and “DB3” in Fig. 2. These DBBs combine information from neighboring layers before sending it to object proposal building blocks (OPBB). In the part of the network that finalizes object detection, the process incorporates a region of interest (ROI) pooling layer and a fully connected (FC) layer.

A key addition to the MS-CNN model is the introduction of deconvolution building blocks (DBB). Each DBB includes extra deconvolution layers and lateral connections, linking feature output layers of different depths to enrich shallow feature maps with deeper semantic details. The proposed CNN model incorporates three DBBs, as depicted in Fig. 2, beginning with a convolution layer (“Conv  $1 \times 1 \times 512$ ”) featuring 512  $1 \times 1$  filters.

The main function of the OPBB is to take feature map outputs from either the DBBs or the Pool6 layer and produce well-defined proposals for subsequent processing by the soft-NMS building block. In this research, there are four OPBBs with a consistent architecture but differing parameters, labeled as “OPBB8,” “OPBB16,” “OPBB32,” and “OPBB64,” as depicted in Fig. 3.

Anchor boxes are pivotal components in regional proposal networks utilized by models like Faster-RCNN and its variants such as MS-CNN. In standard Faster-RCNN, each convolutional filter layer is associated with nine distinct anchor box types. In the baseline MS-CNN, multiple convolutional filter layers within each object proposal building block (OPBB) correspond to specific anchor box types.



**Fig. 3** Object proposal building block

After the object proposal layers, a soft-NMS building block is used to enhance proposals that exhibit substantial overlap. While modern object detection CNN models typically employ the non-maximum suppression (NMS) algorithm, as seen in MS-CNN, the soft-NMS building block is introduced to more effectively handle challenges posed by highly overlapping proposals.

### 3.1 Datasets and Performance Evaluation Metrics

The training is conducted using MATLAB R2021b, utilizing a publicly available dataset from [41]. The dataset comprises 40,000 colored images (resolution:  $227 \times 227$ ), evenly distributed into two subsets.

This research employs the following metrics: Accuracy, Precision, Recall, and F1-measure.

## 4 Results

This section details the outcomes obtained from simulating the network using MATLAB 2021a, and the results are presented in both tabular and graphical formats. Standard performance classification evaluation metrics commonly employed in computer vision applications were utilized for analysis. The dataset was imported into the simulation software and provided as input to the model.

Experimental setup and model implementation in the experiments, the proposed deep neural network underwent training on an NVIDIA GTX 1080 Ti GPU, featuring 3584 CUDA cores and 11 GB GDDR5X memory with a memory bandwidth of 484 GB/s.

In the experiment, we initially acquired a pre-trained detector to expedite the process and bypass the need for extensive training time. Next, the datasets were input

**Table 1** Settings for parameters

Parameters	Settings
Input size	[300 300 3]
Mini batch size	8
Initial learning rate	0.001
Penalty threshold	0.5
Warmup period	1000
l2Regularization	0.0005
Max epochs	80
Verbose frequency	50

into the network. To improve network accuracy during training, data augmentation was implemented, which included applying random transformations to the original data. Through data augmentation, we introduced additional diversity to the training data without the necessity of expanding the labeled training samples.

We define the training options according to the parameters outlined (see Table 1).

The model underwent training using the specified parameter settings and environmental configuration. The data were segregated into training, validation, and test sets using the function `splitEachLabel`, which divided the image datastore into two new data stores, encompassing normal and crack images. Given the dataset's substantial size, the training time was a consideration, prompting the adoption of a low training image ratio, such as 0.1, resulting in a small number of images.

Table 2 presents the results achieved by the proposed system, showing precision and recall values obtained from simulation results across multiple iterations. Accuracy, in certain contexts, can be misleading. Precision and recall offer additional insights into the reliability of the accuracy demonstrated for a specific problem. The F1-Measure provides a consolidated score that addresses concerns related to both precision and recall in a single numerical metric. In binary classification statistical analysis, the F1-score (or F1-measure) evaluates the accuracy of a test by considering both precision and recall. Precision assesses correctly identified positive results divided by all positive results (including misidentifications), while recall measures correctly identified positive results divided by all samples that should have been identified as positive. A higher F1-score typically indicates higher confidence in the classification performance.

**Table 2** The detection rate of the proposed model based on four different OPBB architectures

Performance metric	OPBB8	OPBB16	OPBB32	OPBB64
Precision	0.9895	0.9897	0.9972	0.9985
Accuracy	0.9970	0.9974	0.9984	0.9989
Recall	0.9839	0.9862	0.9892	0.9942
F <sub>1</sub> -score	0.9748	0.9756	0.9797	0.9821

**Table 3** Detection rate of the proposed model against the existing model based on four different OPBBs architectures

Performance metric	Proposed model				Existing model [40]
	OPBB8	OPBB16	OPBB32	OPBB64	
Precision	0.9895	0.9897	0.9972	0.9985	0.9955
Accuracy	0.9970	0.9974	0.9984	0.9989	0.9870
Recall	0.9839	0.9862	0.9892	0.9942	0.9339
F <sub>1</sub> -score	0.9748	0.9756	0.9797	0.9821	0.9548

As shown in Table 3, the proposed model exhibits outstanding performance in terms of accuracy, precision, recall, and F1 score. The following sections will provide an in-depth discussion, offering a thorough analysis and evaluation of these results compared to an existing model.

## 5 Discussions

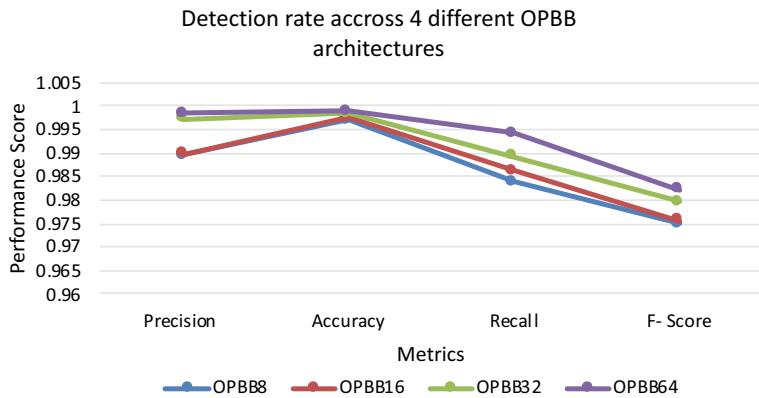
After conducting experiments on the dataset, this section discusses the results of the proposed model on road crack datasets, comparing it with other state-of-the-art techniques. The presentation of results is divided into two subsections. First, a comprehensive evaluation of the proposed model on road crack datasets is provided. Second, the classification performance of the proposed study is benchmarked against other state-of-the-art approaches. To evaluate the detection performance of the proposed approach, accuracy, precision, and recall were computed based on the parameter settings specified in Table 1.

According to Table 3, the proposed model exhibited commendable detection performance across the four different OPBB architectures evaluated in the experiment. Notably, it was observed that the proposed model achieved the highest precision, accuracy, recall, and F1 score in the case of OPBB64. This outcome is more comprehensively illustrated in Fig. 4.

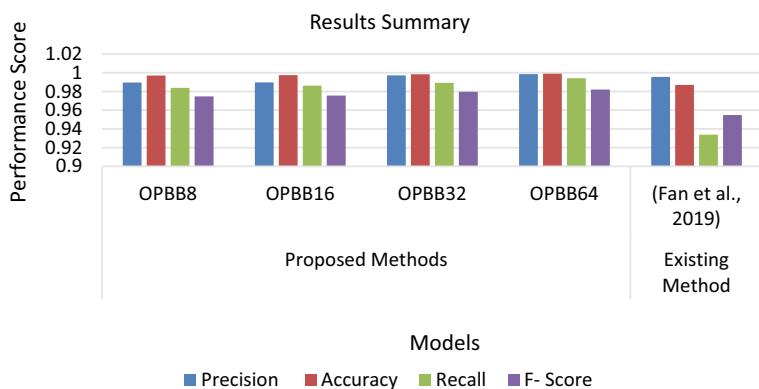
Examining Fig. 4 reveals that at OPBB64, the proposed model achieves the highest detection rates: 0.9985 for precision, 0.9989 for accuracy, 0.9942 for recall, and 0.9821 for F1-Score. Additionally, an observed trend indicates that as the number of OPBBs increases from 8 to 64, the overall detection rate of the model increases, and vice versa.

Moreover, Table 4 displays the results obtained by the proposed model when compared with the existing model.

From Fig. 5, it is evident that the proposed model, featuring four distinct OPBB architectures, consistently achieves top rankings in terms of accuracy, precision, recall, and F1 measure. Notably, the existing model exhibits a slight superiority over the proposed model, particularly in precision rates for OPBB8 and OPBB16.



**Fig. 4** Trend in road crack detection rate for the proposed model base on four different OPBBs



**Fig. 5** Results summary across the four-performance metrics used in this study

## 6 Conclusion

A framework based on multi-scale convolutional architecture for road crack detection was proposed in this study. The framework was developed to enhance the DCNN frameworks, which use soft NMS, specifically addressing object occlusion challenges discovered in the reviewed studies. The developed deconvolution CNN model for crack detection at low feature map scales, coupled with soft NMS, effectively tackles object occlusion challenges and provides meaningful segmentation of positive images. Extensive experiments on the road crack dataset validate the proposed CNN enhancements, showcasing robust detection performance across various OPBB architectures.

Thus, the proposed framework solidifies its overall superiority by achieving higher values in precision, recall, and F1-score, demonstrating its effectiveness in classifying road cracks within images subjected to various occlusions and environmental conditions, surpassing the performance of the existing model. The overall results indicate the proposed deep learning model's consistent first-place rankings across all OPBB cases compared to the benchmark algorithms. Furthermore, the proposed deep learning approach consistently attains near-perfect values, close to 100%, in accuracy, recall, precision, and F1-measure.

This approach effectively addresses challenges related to the significant object scale variation in road crack images, coupled with varying object occlusion, successfully segmenting positive images into semantically meaningful regions, specifically cracks and road surfaces. As a result, it enhances the conventional CNN model's performance in detecting cracks on roads and highways.

In the future, we recommend that researchers evaluate the computational time of the proposed enhancement compared to the existing model as an index for assessing model quality. Additionally, further investigations could explore additional CNN models and enhancement to advance object detection, contributing to safer and more intelligent transportation systems.

**Acknowledgements** The publication of this work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia, under Project GRANTA,262.

## References

1. Garber NJ, Hoel LA, Sarkar R (2009) Traffic and highway engineering
2. Wormi NM et al (2022) Deeper architecture for brain age prediction based on MRI images using transfer learning technique. *Procedia Comput Sci* 212:441–453
3. Li J, Zhao X, Li H (2019) Method for detecting road pavement damage based on deep learning. In: *Health monitoring of structural and biological systems XIII*. SPIE
4. Ayoub S, Gulzar Y, Rustamov J, Jabbari A, Reegu FA, Turaev S (2023) Adversarial approaches to tackle imbalanced data in machine learning. *Sustain* 15(9):7097
5. Yusuf S et al (2023) A review on prediction of Covid-19 cases using machine learning for effective public health management. *ATBU J Sci, Technol Educ* 11(2):202–215
6. Mazzoli A, Monosi S, Plesscia ES (2015) Evaluation of the early-age-shrinkage of fiber reinforced concrete (FRC) using image analysis methods. *Constr Build Mater* 101:596–601
7. Hsieh Y-A, Tsai YJ (2020) Machine learning for crack detection: review and model performance comparison. *J Comput Civ Eng* 34(5):04020038
8. Yunusa AA et al (2023) An adversarial examples against deep learning-based network intrusion detection system: a review. *ATBU J Sci, Technol Educ* 11(2):189–204
9. Jiang W et al (2020) HDCB-Net: a neural network with the hybrid dilated convolution for pixel-level crack detection on concrete bridges. *IEEE Trans Industr Inf* 17(8):5485–5494
10. Umar A et al (2023) An ensemble learning approach to software requirement classification with recursive feature elimination and balance bagging classifier. *ATBU J Sci, Technol Educ* 11(4):9–23
11. Zhang L et al (2016) Road crack detection using deep convolutional neural network. In: *2016 IEEE international conference on image processing (ICIP)*. IEEE

12. Bala Z et al (2022) Transfer learning approach for malware images classification on android devices using deep convolutional neural network. *Procedia Comput Sci* 212:429–440
13. Salman M et al (2013) Pavement crack detection using the Gabor filter. In: 16th international IEEE conference on intelligent transportation systems (ITSC 2013). IEEE
14. Malik I, Ahmed M, Gulzar Y, Baba SH, Mir MS, Soomro AB, Sultan A, Elwasila O (2023) Estimation of the extent of the vulnerability of agriculture to climate change using analytical and deep-learning methods: a case study in Jammu, Kashmir, and Ladakh. *Sustain* 15(14):11465
15. Majid M, Gulzar Y, Ayoub S, Khan F, Reegu FA, Mir MS, Jaziri W, Soomro AB (2023) Enhanced transfer learning strategies for effective kidney tumor classification with CT imaging. *Int J Adv Comput Sci Appl* 14:2023
16. Makaremi M, Razmjooy N, Ramezani M (2018) A new method for detecting texture defects based on modified local binary pattern. *SIViP* 12:1395–1401
17. Idris IYU., Ya'u BI, Ali U (2023) Crackdet: an improved deep learning framework base on multi-scale convolutional architecture for detecting road cracks. *ATBU J Sci, Technol Educ* 11(2):116–124
18. Zou Q et al (2012) CrackTree: automatic crack detection from pavement images. *Pattern Recogn Lett* 33(3):227–238
19. Oliveira H, Correia PL (2012) Automatic road crack detection and characterization. *IEEE Trans Intell Transp Syst* 14(1):155–168
20. Ayoub S, Gulzar Y, Reegu FA, Turaev S (2022) Generating image captions using bahdanau attention mechanism and transfer learning. *Symmetry* 14(12):2681
21. Oliveira H, Correia PL (2014) CrackIT—an image processing toolbox for crack detection and characterization. In: 2014 IEEE international conference on image processing (ICIP). IEEE
22. Ahmad A, Lawal MA An optimized deep learning method for software defect prediction using whale optimization algorithm
23. Gital DD et al (2023) An improved model for electricity load and price prediction using hybrid deep learning algorithms: a comprehensive review. *ATBU J Sci, Technol Educ* 11(3):103–118
24. Elghaish F et al (2022) Developing a new deep learning CNN model to detect and classify highway cracks. *J Eng, Des Technol* 20(4):993–1014
25. Aihong AA et al (2023) An optimized deep learning method for software defect prediction using whale optimization algorithm (WOA): a review. *ATBU J Sci, Technol Educ* 11(3):159–174
26. Gulzar Y, Alwan AA, Abdullah RM, Abualkishik AZ, Oumrani M (2023) OCA: ordered clustering-based algorithm for e-commerce recommendation system. *Sustain* 15(4):2947
27. Gulzar Y (2024) Enhancing soybean classification with modified inception model: a transfer learning approach. *Emir J Food Agric* 36:1–9
28. Mehmood A, Gulzar Y, Ilyas QM, Jabbari A, Ahmad M, Iqbal S (2023) SBXception: a shallower and broader xception architecture for efficient classification of skin lesions. *Cancers* 15(14):3604
29. Alkanan M, Gulzar Y (2024) Enhanced corn seed disease classification: leveraging MobileNetV2 with feature augmentation and transfer learning. *Front Appl Math Stat* 9:1320177
30. Amri E, Gulzar Y, Yeafi A, Jendoubi S, Dhawi F, Mir MS (2024) Advancing automatic plant classification system in Saudi Arabia: introducing a novel dataset and ensemble deep learning approach. *Model Earth Syst Environ* 10(2):2693–2709
31. Liu Y et al (2020) Deep network for road damage detection. In: 2020 IEEE international conference on big data (Big Data). IEEE
32. Haciefendioğlu K, Başağa HB (2022) Concrete road crack detection using deep learning-based faster R-CNN method. *Iran J Sci Technol, Trans Civ Eng* 46(2):1621–1633
33. Hoang ND et al (2022) A novel approach for detection of pavement crack and sealed crack using image processing and salp swarm algorithm optimized machine learning. *Adv Civ Eng*
34. Du F-J, Jiao S-J (2022) Improvement of lightweight convolutional neural network model based on YOLO algorithm and its research in pavement defect detection. *Sensors* 22(9):3537
35. Hu GX et al (2021) Pavement crack detection method based on deep learning models. *Wirel Commun Mob Comput* 2021:1–13

36. Guan J et al (2021) Automated pixel-level pavement distress detection based on stereo vision and deep learning. *Autom Constr* 129:103788
37. Kanaeva I, Ivanova JA (2021) Road pavement crack detection using deep learning with synthetic data. In: IOP conference series: materials science and engineering. IOP Publishing
38. Praticò FG et al (2020) Detection and monitoring of bottom-up cracks in road pavement using a machine-learning approach. *Algorithms* 13(4):81
39. Feng X et al (2020) Pavement crack detection and segmentation method based on improved deep learning fusion model. *Math Probl Eng* 2020:1–22
40. Fan R et al (2019) Road crack detection using deep convolutional neural network and adaptive thresholding. In: 2019 IEEE intelligent vehicles symposium (IV). IEEE
41. Dataset link: [https://github.com/ruirangerfan/road\\_crack\\_detection\\_net.git](https://github.com/ruirangerfan/road_crack_detection_net.git). Created by researchers from Middle East Technical University. Accessed 15 Apr 2024

# Knowledge Graph Relation Learning Using GAN-BERT



Neelam Jain  and Krupa Mehta

**Abstract** Massive knowledge graphs (KGs) are increasingly significant in modern information systems. Previous research on knowledge graph completion must gather sufficient training instances for new introduced relations in order to increase the coverage of KGs. This study examines an innovative approach to the scarce labelling problem. The model uses text descriptions to learn semantic elements of newly added relations, allowing us to recognise facts even when there are few examples available. To learn relations based on text, we utilised a GAN-BERT model, which is a modified form of BERT. Four distinct datasets are employed for this. Experiments demonstrate that using GAN-BERT decreases the demand for annotated instances and improves performance in various relation prediction tasks.

**Keywords** BERT · Relation prediction · Knowledge graph

## 1 Introduction

A knowledge graph is a systematic approach of representing information as a set of entities (nodes) and their connections (edges). These graphs are used to model and organise many forms of data, including facts, concepts and entity connections. Figure 1 shows a knowledge graph. Edges connecting entity nodes denote relations. The relations are organised as triplets (head, relation, tail), with Harry Potter as the head, Gryffindor as the tail and belongsToHouse as the relation.

The graph-structured knowledge base is a valuable resource that can be used for search engines, recommendation systems and question answering. The problem

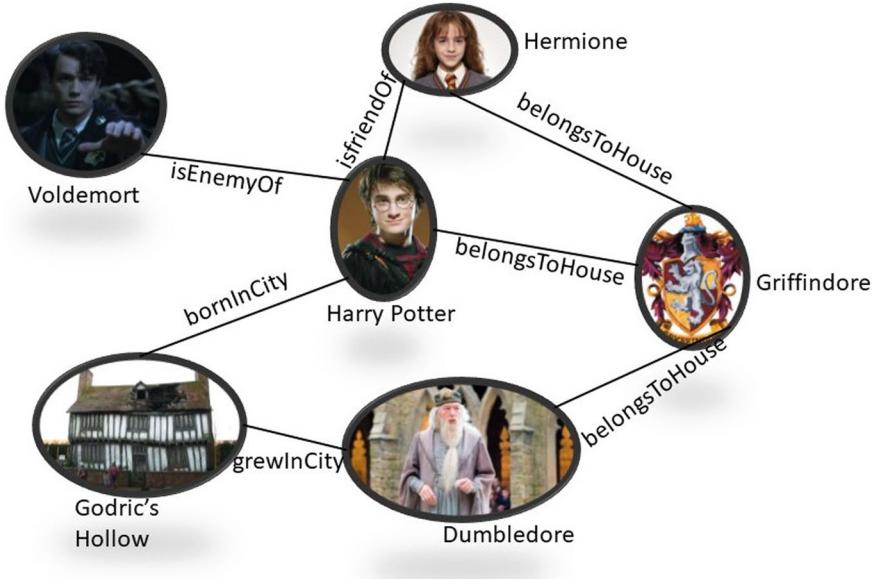
---

N. Jain ()

SVKM's Mithibai College of Arts, Chauhan Institute of Science & Amrutben Jivanlal College of Commerce And Economics, Mumbai, India  
e-mail: [neelamdoshi@gmail.com](mailto:neelamdoshi@gmail.com)

K. Mehta

Faculty of Computer Application and IT, GLS University, Ahmedabad, Gujarat, India  
e-mail: [krupa.mehta@glsuniversity.ac.in](mailto:krupa.mehta@glsuniversity.ac.in)



**Fig. 1** Knowledge graph subset

with current knowledge graph completion consists of predicting or inferring missing links or edges in a knowledge graph. Numerous research [1–3] have demonstrated remarkable success in automatically identifying and filling in the gaps in the existing relations in order to address this issue. These techniques use triplet information without considering the semantics associated with them. While several KGE models have demonstrated significant effectiveness in automatically identifying and completing gaps in the relational data that already exists, they are unable to identify relations that lack sufficient training data or previously unseen [4]. The unseen information refers to a new triple combination rather than a new object or relation. Also, there is still a non-negligible constraint for recently added relations, and getting sufficient training instances for each new relation becomes a more difficult solution.

The ability to identify new relations just by reading their text descriptions is a compelling aspect of human comprehension that becomes apparent after a given amount of experience. Thus, the semantic characteristics of new triplets can be expressed by their textual descriptions, rather than learning from training triplets. Furthermore, textual descriptions are essential for large-scale recognition tasks since they are easily accessible via dictionaries, encyclopaedia articles, and other internet resources and include extensive and clear information.

The article presents a GAN-BERT [5] architecture-based approach for relation learning from textual descriptions. The study aims to use text descriptions to establish appropriate relations with a smaller training set. The objective is to detect knowledge graph relations based on textual descriptions and statements constructed with head and tail. Use of GAN-BERT improves relation learning with minimal annotated data

and unlabeled sources. This study uses a variety of training set sizes to show that the GAN-BERT can deliver good results with relatively little annotated data.

The following is a brief summary of this research's contributions:

- Based on textual description, the GAN-BERT architecture is extended to estimate the likelihood of a new knowledge graph relations.
- The model is evaluated on four benchmark datasets: Yago3-10[6], NELL [7], WN18RR [3] and FB15k-237 [8]. As opposed to ranking-based evaluation, which is typical in other knowledge graph embeddings, classification-based evaluation [9] is carried out for performance.
- Testing triplets, which are not present during training, are used to assess performance.

The remainder of the paper is organised as follows. Related work on knowledge graph embeddings is covered in Sect. 2. Section 3 describes the implementation of the SS-GAN [10] based GAN-BERT [5] model. The proposed GAN-BERT-based relation learning model is presented in Sect. 4. Section 5 presents promising experimental results for the benchmark Yago3-10, NELL, WN18RR and FB15k-237 datasets. Section 6 has concluding remarks.

## 2 Related Work

Knowledge Graph Embedding approaches project knowledge graph elements to a continuous vector space to understand their representation and relationships. Several studies have attempted to predict missing connections in knowledge graphs. Knowledge graph embedding techniques fall into three categories: translational-distance-based, tensor decomposition and neural network models. Translational distance models use distance as a scoring variable and relation as the head-to-tail translation [1, 11, 12]. Tensor decomposition models use tensor products to represent complex relationships [2, 13]. Neural network-based models [3, 14, 15] facilitate self-learning. These models have been found to work well with previously encountered triplets, but they fail when the triplet pair is unknown and in the raw text format. KBGAN [16] employs adversarial training to improve discriminators by picking high-quality negative samples. However, it mostly focuses on predicting existing relations. Representation learning is now the most extensively used method for modelling knowledge graph information [17]. Text-based approaches generate intermediate representations of semantics based on unstructured text data available online [18]. It proposes a method for suppressing noise in text while also learning a function to match text and visual elements. This strategy is intended to reduce the cost of manually collecting semantic features while also enhancing the effectiveness of zero-shot learning models in noisy situations. However, it does not investigate the possible advantages of combining its noise suppression strategy with different semantic sources or zero-shot learning methods.

Training supervised Relation extraction models takes time due to the need for extensive annotated data [19–21]. As a result, [22] proposed DS for automatically labelling data. It invariably comes with an issue with incorrect labelling. The OpenRE [23] approach seeks to discover the word or sequence of words within a sentence that expresses a relationship between two entities. GRL [24] suggests utilising a combination of GANs and reinforcement learning approaches to fill in missing relationships in a knowledge graph but fails to work on unseen relations. Reference [25] addresses relation learning with GANs for textual description. In order to predict links using bag-of-words method, it attempts to employ neighbour encoder. As a result, it cannot take advantage of the BERT’s [26] contextualised word representation.

### 3 Semi-supervised GAN (SS-GAN) and GAN-BERT

Modern Transformer-based architectures, like BERT [26], do remarkably well in a variety of NLP tasks. Nevertheless, most of the benchmark models require a substantial amount of labelled data. While it can be costly and time-consuming to obtain high-quality labelled data in many real-world circumstances, it is generally easy to gather unlabeled examples that characterise the target task. Through the use of a Semi-Supervised Generative Adversarial Learning approach [10], GAN-BERT [5] makes semi-supervised learning possible in BERT-based architectures. In a way, SS-GANs [27] is semi-supervised learning using a GAN framework. “True” instances are classified in one of the target ( $1, \dots, k$ ) classes, and the generated samples are classified into the  $k + 1$  class. The discriminator is trained over a  $(k + 1)$ -class objective. Let  $\mathbb{D}$  and  $\mathbb{G}$  be the discriminator and generator, respectively, while  $p_d$  pd and  $p_g$  signify the real data distribution and generated examples. The objective of  $\mathbb{D}$  is expanded to build a semi-supervised  $k$ -class classifier. Let us define  $p_m(\hat{y} = y|x, y = k + 1)$  the probability given by the model m that a generic example  $x$  is related with the fake class and  $p_m(\hat{y} = y|x, y \in (1, \dots, k))$  that  $x$  is regarded real, therefore associated to one of the target classes. The loss function of  $\mathbb{D}$  is  $L_{\mathbb{D}} = L_{\mathbb{D}sup} + L_{\mathbb{D}unsup}$  where

$$L_{\mathbb{D}sup} = -\mathbb{E}_{x,y \sim p_d} \log[p_m(p_m(\hat{y} = y|x, y \in (1, \dots, k)))] \quad (1)$$

$$\begin{aligned} L_{\mathbb{D}unsup} = & -\mathbb{E}_{x,y \sim p_d} \log[1 - p_m(p_m(\hat{y} = y|x, y = k + 1))] \\ & -\mathbb{E}_{x \sim \mathbb{G}} \log[p_m(\hat{y} = y|x, y = k + 1)] \end{aligned} \quad (2)$$

$L_{\mathbb{D}sup}$  calculates the error in allocating the wrong class to a real-world sample from the original  $k$  categories.  $L_{\mathbb{D}unsup}$  assesses the error in incorrectly identifying a real, unlabelled case as fake while failing to identify a false example. In addition,  $\mathbb{G}$  is expected to offer examples that are similar to those taken from the real distribution. As per [27],  $\mathbb{G}$  ought to produce data that roughly approximates the distribution of actual data. In a nutshell, the average example produced by  $\mathbb{G}$  in a batch should be equivalent to the actual prototype example. Let’s use the notation  $f(x)$  to represent

the activation on one of  $\mathbb{D}$ 's intermediary layers. The feature matching loss for  $\mathbb{G}$  is thus defined as

$$L_{\mathbb{G} \text{ featurematching}} = \|\mathbb{E}_{x \sim p_d} f(x) - \mathbb{E}_{x \sim \mathbb{G}} f(x)\|_2^2 \quad (3)$$

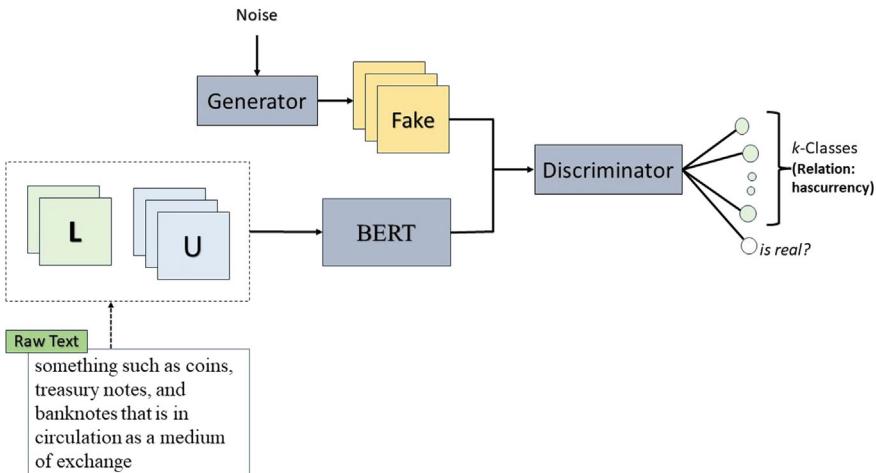
that is, the generator should generate cases in which the intermediate representations presented to  $\mathbb{D}$  are substantially comparable to the real ones. The  $\mathbb{G}$  loss takes into account the error caused by fraudulent examples that were accurately identified by  $\mathbb{D}$ , that is,

$$L_{\mathbb{G} \text{ unsup}} = -\mathbb{E}_{x \sim \mathbb{G}} \log[1 - p_m(\hat{y} = y|x, y = k+1)] \quad (4)$$

The  $\mathbb{G}$  loss is  $L_{\mathbb{G}} = L_{\mathbb{G} \text{ featurematching}} + L_{\mathbb{G} \text{ unsup}}$ . GAN-BERT [5] enhances the pre-trained BERT model with two additional elements for fine-tuning: (i) task-specific layers, like in standard BERT fine-tuning; (ii) SS-GAN layers for semi-supervised learning.

## 4 Proposed Model

The knowledge graph is constantly evolving, with new relations uncovered on a regular basis. In this situation, it is quite usual to not have an adequate training set. The semi-supervised GAN-BERT model facilitates relation learning with a smaller training set. Figure 2 illustrates the structure of GAN-BERT for relation learning. As shown in Fig. 2 instead of using the relation name, the textual representation of relations is joined with the triplet head and tail to make a sentence. Since Textual



**Fig. 2** Structure of GAN-BERT

descriptions of relation provide comprehensive and understandable information and are readily available through dictionaries, encyclopaedia articles and other online sites. These sentences also enhance the utilisation of BERT [26] contextualization. A small number of labelled and unlabeled statements are provided to the GAN-BERT Model for training purposes. Here, “labels” refers to the relation name. The model assigns the sentences to the appropriate class or relation name, or it classifies them as fake. For this purpose, at least one relation with its textual description must be present in the training set. The technique employs a few-shot scenario, requiring only a few instances to train for unseen relations.

The original BERT implementation is extended in pyTorch to provide GAN-BERT for relational learning. Generator is a multi-layer perceptron (MLP) with a leaky-relu function to activate one hidden layer. Generator inputs are noise vectors derived belonging to the Gaussian distribution  $N(0, 1)$ . The noise vectors run through the Multi-layer perceptron, resulting in vectors with 768 dimensions that we employ as fake examples in GAN-BERT framework. Discriminator is also an MLP, with one hidden layer activated by the leaky-relu function and a softmax layer for the ultimate prediction. The comparison of several state-of-the-art KGE models is not included since popular KGE models rely on scoring functions, whereas this technique uses classification [9].

## 5 Experiments and Results

This section evaluates GAN-BERT’s performance on relation learning tasks with varying training settings, such as the quantity of instances and relations. The model is tested using 37 relations from the Yago3-10 dataset [6], 117 from the NELL [7] dataset, 11 from the WN18RR [3] dataset and 237 from the FB15k-237 [8] dataset. The meaning of the relation is defined by the textual description of the each relation. Table 1 shows relation instances with labels and descriptions in text. Table 2 summarises the dataset information.

The training sets for the Yago3-10, WN18RR and FB15K-237 datasets are generated at random, with 111, 33 and 711 triplets per relation, respectively. These triplets are constructed by combining relevant entities from the Yago3-10 training set with the relation’s textual description. The purpose of this research is to categorise relations based on text descriptions with distinct occurrences of each relation. For testing, 3500 triplets are chosen at random. The training set of the NELL data set contains 468 randomly selected triplets, while the testing set has 42935 randomly selected triplets. The training triplets are the labelled triplets in both datasets. To assess the influence of the size of the unlabeled dataset, various sized triplets are employed in all of the above datasets. 1,000 triplets were randomly selected from Yago3-10, 4,000 triplets from WN18RR, 2,000 triplets from FB15K-237 and 9,000 triplets from NELL as an unlabeled dataset. Only labelled samples are utilised to feed into generative networks. The triplets are represented in the form of sentences, such as “auburn this is another name for this organisation state university.”

**Table 1** Few instances of relation and its textual description

Dataset	Textual description	Relation
Yago3-10	A human offspring son or daughter of any age	hasChild
	An inhabitant of a particular place	isCitizenOf
	Situated in a particular place	isLocatedIn
NELL	Describes the tributaries of a river	Riveremptiesintoriver
	Animals that are hunted and eaten by this animal	Animalpreyson
	To come into possession or ownership of; get as one's own	Acquired

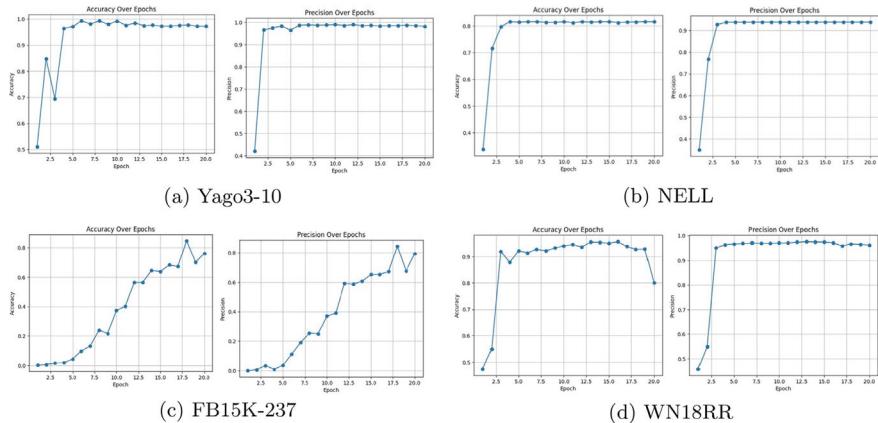
**Table 2** Datasets for evaluation

Dataset	#Relations	#Training triplets	#Testing triplets
Yago3-10	37	111	3500
NELL	117	468	42935
WN18RR	11	33	5000
FB15K-237	237	711	7000

**Table 3** Performance metric of GAN-BERT-based model

Training set	Accuracy	Precision	Recall	F1-score
Yago3-10	0.97	0.98	0.97	0.98
NELL	0.82	0.94	0.81	0.8
WN18RR	0.80	0.96	0.80	0.83
FB15K-237	0.76	0.79	0.76	0.73

The obtained performance measures provide high accuracy with a limited amount of labelled data. It is discovered that given a restricted dataset, as the unlabeled dataset grows, so does the accuracy. When the size of the unlabeled triplet in training set rises, we can attain good accuracy with fewer epochs. Table 3 shows the performance metrics for GAN-BERT-based classification of relations. As can be observed, the model did not perform well on the FB15K-237 dataset when compared to other datasets. Surprisingly, it works well by expanding only the unlabeled dataset. For each of the four datasets, the model has been evaluated on varying sizes of the unlabeled dataset. In every case, it worked well and produced higher accuracy with fewer epochs. It should also be noted that the model's performance is not impacted by variations in testing size. Figure 3 depicts the model's accuracy and precision across different training dataset sizes and epochs. It should be noted that the training phase head and tail values are different from those used in the testing set.



**Fig. 3** Accuracy and precision over epochs for different datasets

## 6 Conclusion

This study utilised GAN-BERT for relational learning in low-resource environments. The study shows that employing text descriptions to learn semantic elements of relationships can improve classification performance in knowledge graphs. The GAN-BERT model shows potential for decreasing the requirement for annotated instances while maintaining high accuracy. Generative adversarial networks can be a good model for unseen relations with limited resources. The model performed effectively even when the training set was significantly smaller than the testing data set. The model's performance suggests that it can further be applied to link prediction in addition to relation prediction.

## References

1. Bordes A, Usunier N, Garcia-Duran A, Weston J, Yakhnenko O (2013) Translating embeddings for modeling multi-relational data. *Adv Neural Inf Process Syst* 26
2. Trouillon T, Welbl J, Riedel S, Gaussier É, Bouchard G (2016) Complex embeddings for simple link prediction. In: International conference on machine learning, PMLR, pp 2071–2080
3. Dettmers T, Minervini P, Stenetorp P, Riedel S (2018) Convolutional 2d knowledge graph embeddings. In: Proceedings of the AAAI conference on artificial intelligence, vol 32
4. Peng C, Xia F, Nasiriparsa M, Osborne F (2023) Knowledge graphs: opportunities and challenges. *Artif Intell Rev* 56(11):13071–13102
5. Croce D, Castellucci G, Basili R (2020) GAN-BERT: Generative adversarial learning for robust text classification with a bunch of labeled examples. In: Jurafsky D, Chai J, Schluter N, Tetreault J (eds) Proceedings of the 58th annual meeting of the association for computational linguistics, (Online). Association for Computational Linguistics, pp 2114–2119
6. Mahdisoltani F, Biega J, Suchanek FM (2013) Yago3: a knowledge base from multilingual wikipedias. In: CIDR

7. Carlson A, Betteridge J, Kisiel B, Settles B, Hruschka E, Mitchell T (2010) Toward an architecture for never-ending language learning. Proceedings of the AAAI conference on artificial intelligence 24:1306–1313
8. Toutanova K, Chen D (2015) Observed versus latent features for knowledge base and text inference. In: Allauzen A, Grefenstette E, Hermann KM, Larochelle H, Yih SW-T (eds) Proceedings of the 3rd workshop on continuous vector space models and their compositionality(Beijing, China). Association for Computational Linguistics, pp 57–66
9. Speranskaya M, Schmitt M, Roth B (2021) Ranking versus classifying: measuring knowledge base completion quality. arXiv preprint [arXiv:2102.06145](https://arxiv.org/abs/2102.06145)
10. Sricharan K, Bala R, Shreve M, Ding H, Saketh K, Sun J (2017) Semi-supervised conditional gans. arXiv preprint [arXiv:1708.05789](https://arxiv.org/abs/1708.05789)
11. Wang Z, Zhang J, Feng J, Chen Z (2014) Knowledge graph embedding by translating on hyperplanes. In: Proceedings of the AAAI conference on artificial intelligence, vol 28
12. Lin Y, Liu Z, Sun M, Liu Y, Zhu X (2015) Learning entity and relation embeddings for knowledge graph completion. In: Twenty-ninth AAAI conference on artificial intelligence
13. Yang B, Yih W-T, He X, Gao J, Deng L (2014) Embedding entities and relations for learning and inference in knowledge bases. arXiv preprint [arXiv:1412.6575](https://arxiv.org/abs/1412.6575)
14. Nguyen DQ, Nguyen TD, Nguyen DQ, Phung D (2017) A novel embedding model for knowledge base completion based on convolutional neural network. arXiv preprint [arXiv:1712.02121](https://arxiv.org/abs/1712.02121)
15. Vashishth S, Sanyal S, Nitin V, Agrawal N, Talukdar P (2020) Interacte: Improving convolution-based knowledge graph embeddings by increasing feature interactions. Proceedings of the AAAI conference on artificial intelligence 34:3009–3016
16. Cai L, Wang WY (2017) Kbgan: adversarial learning for knowledge graph embeddings. arXiv preprint [arXiv:1711.04071](https://arxiv.org/abs/1711.04071)
17. Nickel M, Tresp V, Kriegel H-P (2011) A three-way model for collective learning on multi-relational data. In: Proceedings of the 28th international conference on international conference on machine learning, ICML'11, (Madison, WI, USA). Omnipress, pp 809–816
18. Qiao R, Liu L, Shen C, van den Hengel A (2016) Less is more: zero-shot learning from online textual documents with noise suppression
19. Zelenko D, Aone C, Richardella A (2003) Kernel methods for relation extraction. J Mach Learn Res 3:1083–1106
20. Wang G, Zhang W, Wang R, Zhou Y, Chen X, Zhang W, Zhu H, Chen H (2018) Label-free distant supervision for relation extraction via knowledge graph embedding. In: Proceedings of the 2018 conference on empirical methods in natural language processing, pp 2246–2255
21. Mooney R, Bunescu R (2005) Subsequence kernels for relation extraction. Adv Neural Inf Process Syst 18
22. Mintz M, Bills S, Snow R, Jurafsky D (2009) Distant supervision for relation extraction without labeled data. In: Proceedings of the joint conference of the 47th annual meeting of the ACL and the 4th international joint conference on natural language processing of the AFNLP, pp 1003–1011
23. Yu D, Huang L, Ji H (2017) Open relation extraction and grounding. In: Kondrak G, Watanabe T (eds) Proceedings (volume 1: long papers) (Taipei, Taiwan). Asian Federation of Natural Language Processing pp 854–864
24. Wang Q, Ji Y, Hao Y, Cao J (2020) Grl: knowledge graph completion with gan-based reinforcement learning. Knowl Based Syst 209:106421
25. Qin P, Wang X, Chen W, Zhang C, Xu W, Wang WY (2020) Generative adversarial zero-shot relational learning for knowledge graphs
26. Devlin J, Chang M-W, Lee K, Toutanova K (2019) Bert: pre-training of deep bidirectional transformers for language understanding
27. Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X (2016) Improved techniques for training gans

# SightAssist: A Multi-facility Machine Learning Approach for Empowering the Visually Impaired



Saikat Bandopadhyay, Jhalak Dutta, Smita Das, Soumyajit Datta, Debaditya Ghosh, Rohit Kumar Dey, and Jeet Nandigrami

**Abstract** Blindness or visual impairment ranks among the top ten disabilities affecting both men and women, impacting over 35 million individuals of all ages in India. Access to visual information is crucial for enhancing the independence and safety of blind and visually impaired individuals. There exists a compelling imperative to develop assistive technologies that can significantly improve their quality of life. This paper presents SightAssist, a multilingual software solution designed to address the challenges faced by the visually impaired community in India. SightAssist utilizes a camera-based approach to serve as a digital prosthetic eye and virtual assistant, capable of currency detection, facial recognition, and providing various functionalities to assist with daily activities. Initially, a custom-labeled dataset has been prepared using a pre-trained YOLO model. Gradually, dataset pre-processing, normalization, and semantic segmentation are employed to optimize model performance. Additionally, voice-based accessibility features have been integrated using the Web Speech API, enabling users to interact with the application through voice commands. The experimental results show the successful development of SightAssist and its potential to significantly improve the lives of visually impaired people.

**Keywords** Assistive technology · Computer vision · Supervised machine learning · Virtual assistant

---

S. Bandopadhyay

Department of Artificial Intelligence and Machine Learning, Netaji Subhash Engineering College, Kolkata, India

J. Dutta (✉) · R. K. Dey

Department of Computer Science and Engineering, Heritage Institute of Technology, Kolkata, India  
e-mail: [jhalak.dutta@heritageit.edu](mailto:jhalak.dutta@heritageit.edu)

S. Das

Department of Computer Science and Engineering, National Institute of Technology, Agartala, India

S. Datta · D. Ghosh · J. Nandigrami

Department of Artificial Intelligence and Machine Learning, Heritage Institute of Technology, Kolkata, India

## 1 Introduction

According to the Indian Journal of Ophthalmology [1], published in June 2022, approximately 4.95 million people in India are blind (0.36% of the total population), with an additional 35 million experiencing visual impairment (2.55%). The country also has an estimated 0.24 million blind children. Vision impairment is ranked among the top ten disabilities affecting both men and women. Individuals with visual impairments face numerous challenges, including reduced spontaneity and flexibility, limited employment opportunities, difficulty in independently performing household tasks and shopping, as well as significant time and effort required for everyday activities. While new technologies are emerging to assist them in leading more independent lives, many of these innovations face hurdles in self-navigation, leading to social discomfort. Previous technologies such as NAVI (Navigation Assistance for the Visually Impaired) [2], VOICE (which converts images into audio cues) [3], TVS (Tactile Vision System) [3], and Tyflos systems [4] which utilize vibration feedback, have attempted to address these issues. However, they suffer from limitations such as slow operation and impracticality, hindering their widespread adoption and effectiveness.

In this paper, a multilingual software solution SightAssist has been introduced. It is a camera-based assistive tool which can detect the currency, recognize the faces, and serve as a digital prosthetic eye as well as a virtual assistant for the visually impaired community in India. Figure 1 shows the available features in the proposed tool.

### 1.1 Motivation

The core motivation lies in the potential of this work is to make a tangible, positive impact on the lives of million visually impaired individuals. Beyond functional



**Fig. 1** Features of SightAssist

enhancements, the aim is to foster independence, enhance safety, and provide swift access to emergency services, contributing to a more inclusive society. SightAssist is accessible with a camera and an internet connection, making it a powerful tool for enhancing the lives of the blind.

## 1.2 Contribution

The contribution of the proposed model lies in developing a content-based filtering approach that accurately predicts user preferences based on their past interactions with items. User profiles are created from explicit ratings and implicit behaviors, allowing the system to map these preferences to similar content and ensure relevant recommendations. The importance of data preprocessing is emphasized to eliminate irrelevant information, thus improving the efficiency and accuracy of the recommendation system. By overcoming the limitations of collaborative filtering, this content-based approach provides a robust solution for personalized content delivery in various domains, including online research libraries, music streaming services, and on-demand video platforms.

The contributions of this paper are as following:

- Providing multiple facilities such as currency recognition, live location tracking, advanced face recognition, emergency alerts, and multilingual support to empower the visually impaired.
- Extensive analysis of the simulation to provide an assistance to the visually impaired.

The rest of the paper is organized in the following way: Sect. 2 presents a brief literature review of the very recent research works in empowering visually impaired people. Moving forward towards Sect. 3, where the proposed methodology is discussed along with dataset and integration of voice-based accessibility features. Results and discussions are reported in Sect. 4 followed by Sect. 5 that concludes the paper while also outlining future Scopes of work.

## 2 Literature Review

Understanding two basic approaches of object detection, i.e., single-stage detectors vs slower two-stage detectors. A study in [5] examines the progress made in YOLO, draws comparisons with two-stage detectors, and discusses potential avenues for further research.

The authors in [6] look at how object identification has changed over time, emphasizing deep learning developments while recognizing that traditional approaches are still useful in some situations. The literature includes a review of object detection

frameworks, backbone CNNs, common datasets, and assessment measures. A comparative study of the model's performance on the PASCAL VOC and MS COCO datasets wraps off the work.

In [7], a thorough examination on deep learning-based semantic segmentation, collaborative segmentation, and traditional segmentation is presented. The algorithms and approaches are explained in detail, along with their benefits and drawbacks, and their application is examined.

A work in [8] presents a unique connected-component segmentation technique to determine plate boundaries for vehicle plate identification, which is part of an approach to traffic infraction monitoring that shows better performance than that of the Otsu thresholding approach.

A technique by [9] that uses YOLO v5 trained on a pooled dataset of 95 items to improve voice synthesis and detection accuracy for visually impaired people. The output labels of the model are translated into text, which is subsequently played out over a speaker. Pyttsx3 and gTTS, two Python libraries for audio conversion, are also compared in the study.

The suggested system proposed in [10] recognizes things and sends audio alerts to help visually impaired people. It utilizes a Single Shot Detector (SSD) model with MobileNet and TensorFlow Lite to identify items and money notes in real time, both indoors and outdoors.

A study uses a pre-trained Caffemodel framework to transform observed objects into text and the MobileNet SSD approach to translate the text to audio, the model enables real-time object identification for visually impaired users [11].

A project presented in [3] uses YOLO for real-time object detection and e-Speak with gTTS for text-to-speech translation to convert visual information into audio for the visually impaired. Using a Raspberry Pi camera, the system takes pictures, processes them to identify things, and outputs aural feedback via a headset.

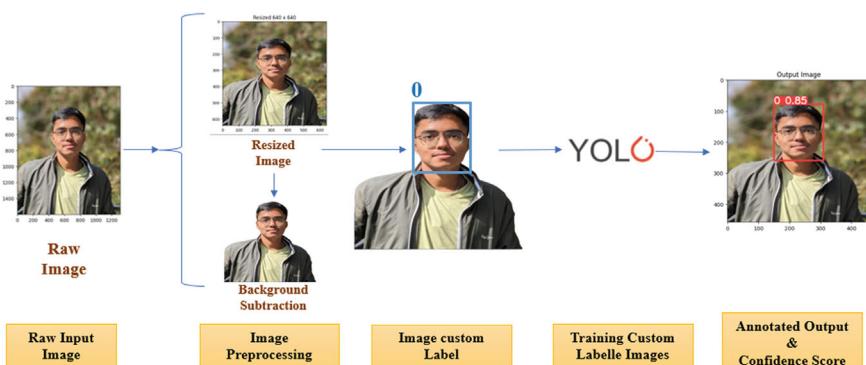
The summarization of all the above-mentioned research work has been tabulated in Table 1.

### 3 Methodology

This section discusses the proposed method for developing the SightAssist application using custom-labeled dataset and pre-trained YOLO model. First, the dataset, pre-processing and its annotation will be discussed. Then, the preprocessing method of the dataset will be explained. In the end, a YOLO model will be introduced for identification of various objects. The Work Flow diagram of the proposed methodology is given in Fig. 2. The raw input images from the collected data set are initially pre-processed based on resizing and background segmentation. Subsequently, a custom label is put on the image to represent the image type followed by training the custom images. Finally, annotated output for image with confidence score is displayed.

**Table 1** Summary of related literature

Paper	Focus	Methodology	Dataset used	Model/framework	Performance/results
[8]	Vehicle number plate recognition	Object detection	Custom dataset	Open-source libraries	Promising outcomes compared to Otsu method for vehicle plate recognition
[7]	Image segmentation	Review	N/A	Various	Segmentation fundamentals, model comparisons, practical applications
[6]	Object detection models, evaluation	Review	PASCAL VOC, MS COCO	Various	Comparison of object detection algorithms, research challenges, performance on datasets
[5]	Single-stage object detection	Review	PASCAL VOC, MS COCO	YOLO	Architectural improvements, performance statistics, comparative analysis, future research directions
[3]	Object detection for visually impaired	Object detection, speech synthesis	N/A	YOLO	Real-time object detection, speech signal conversion
[10]	Assistive technology	Object detection	Custom dataset	MobileNet	Real-time object recognition, suitable for indoor and outdoor environments
[9]	Assistive technology	Object detection, speech synthesis	Custom dataset, MS COCO	YOLO v5	Improved accuracy in detection and speech generation, comparison of audio conversion libraries
[11]	Assistive technology	Object detection, text-to-speech conversion	Pre-trained dataset	MobileNet SSD, Caffemodel	Real-time object detection, text-to-speech conversion

**Fig. 2** Workflow Diagram

**Table 2** Dataset distribution

Image type	Image quantity
Family and friends	120
Currencies	70
Bikes	80
Cycles	80
Roadside individuals	200

### 3.1 Dataset

The dataset includes images taken from camera and various smartphones. These images mainly include family and friends, different denominations of Indian currencies, bikes, cycles, and other random roadside individuals engaged in various task like walking, standing, eating, drinking, etc. Total of 560 images were captured. The distribution of the dataset is given in Table 2.

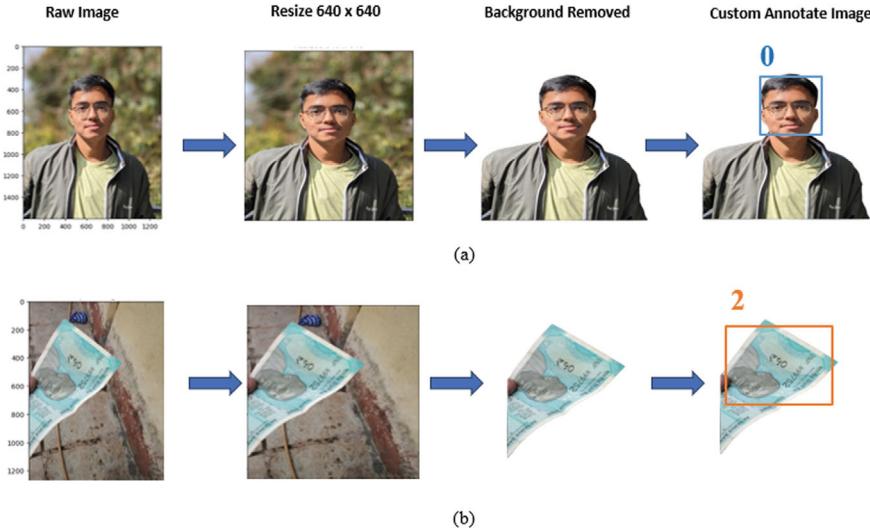
### 3.2 Data Pre-processing

#### 3.2.1 Resizing

The original image are taken from various devices. Images generated consist of different pixel densities and sizes. Thus, pre-processing of all these custom images is highly necessary. Resizing is the first pre-processing step, here all the images are captured and resized to  $640 \times 640$  dimensions. This ensured consistent input for the YOLO classification model that improves computational efficiency. Data normalization and bounding box scaling are crucial pre-processing steps in object detection pipelines, impacting model performance and generalizability. In this model, utilizing YOLOv5 for currency recognition, face recognition, and road assistance, the following strategies were employed:

#### 3.2.2 Normalization

After resizing, it has been normalized the pixel values to the range (0, 1) using minimax normalization technique. This standardization is essential for deep learning models, as it contributes to stable learning, improving convergence, enhancing generalization, and making optimization more manageable.



**Fig. 3** Image segmentation with background removal

### 3.2.3 Data Segmentation

Image segmentation is a technique in Computer Vision that partitions a digital image into discrete groups of pixels. Among various techniques, semantic segmentation has been applied in this proposed approach [12, 13]. The important aspects of semantic segmentation is to accurately identify objects in the image and assign meaningful labels to each pixel. In this article, the semantic segmentation has been explicitly to perform background removal. It is used to outline and exclude backgrounds wherever relevant. Figure 3a and b displays the image segmentation with background removal.

### 3.2.4 Data Annotation

The used dataset was meticulously annotated using bounding boxes to delineate objects of interest. For example, for currency note detection, the bounding boxes around each denomination are annotated. Similarly, for facial recognition, the facial regions within images has been marked. These annotations served as ground truth labels for training of the object detection models [14]. Leveraging tools like Roboflow, has efficiently managed the used dataset by organizing annotations, images, and metadata [15]. Roboflow's interface was utilized to visualize annotations, correct errors, and export datasets in formats compatible with the model YOLOv5. Fig. 3 shows the facial image and currency image annotation in Fig. 3a and b, respectively. Each individual faces are annotate with labeled name starting from 0, 1, ..., n. Similarly, for currencies. Each numbered labeled is mapped with their

corresponding names and the metadata are created. For example, in Fig. 3b currency of denomination amount Rs 50 has been labeled as 2 and denomination amount Rs 10 labeled as 0.

### **3.3 Model Selection and Training**

The proposed project utilizes the YOLO algorithm that provides real-time object detection. Further, used gTTS modules for text to speech [16] and web speech API [17] (javascript API) for voice integration that gives speech on identifying an object.

#### **3.3.1 YOLO V5 for Object Detection**

The You Only Look Once(YOLO) approach to object identification revolutionizes the field by swiftly and accurately detecting objects within images, effectively partitioning the image into a grid and assigning each cell the task of identifying objects within its boundaries [18]. One of the most important features of YOLO v5 is its simplicity in retraining on custom datasets. This characteristic helps the proposed model to work specific requirements with ease, i.e., identifying family and friends and currency of different denominations.

In terms of optimization, YOLO v5 represents a significant advantage from its predecessors. Notably, it achieves a substantial reduction in storage requirements, utilizing approximately 90% less space compared to YOLO v4, while maintaining comparable accuracy [19]. By utilizing the capabilities of this advanced deep neural network, notable object detection results have been achieved with unparalleled speed and accuracy, making YOLO v5 a compelling choice for a wide array of computer vision tasks.

### **3.4 Integration of Voice-Based Accessibility Features**

Voice-based accessibility features have been integrated into the proposed model, which serves as a comprehensive accessibility assistant for visually impaired individuals. Since it is impossible for them to type and scroll through the screen. Thus, developing a user interface (UI) that primarily revolves around a single-button interface for seamless interaction via voice commands. Moreover, the proposed model is designed to be multilingual, ensuring accessibility to a diverse user base.

The Web Speech API is a JavaScript API that enables web developers to incorporate speech recognition and synthesis capabilities into their web applications. It allows users to interact with web applications through speech, which can be particularly useful for accessibility purposes or for creating hands-free user interfaces. Thus, using the web Speech API on the front end to ensure the users can issue simple,

intuitive commands such as “Start road assist mode”, “Who is this person”, or “What is the value of this currency” to initiate specific functionalities tailored to their needs [20]. The Web Speech API consists of two main components.

**Speech Recognition:** This component allows web applications to recognize speech input from the user. Developers can use this feature to transcribe spoken words into text, enabling users to control their applications using voice commands.

**Speech Synthesis:** This component allows web applications to generate synthetic speech from text input. Developers can use this feature to provide auditory feedback to users or to create voice-enabled interfaces where the application responds to user input with spoken responses.

Upon receiving voice commands, the proposed model offers several solutions to address the user’s requirements promptly which are as follows:

1. **Facial Recognition:** Issuing the command “Who is this person” prompts the application to utilize its facial recognition capabilities to identify individuals, providing valuable contextual information to the user.
2. **Currency Value Detection:** Upon receiving the command “What is the value of this currency”, the proposed model employs computer vision algorithms to detect and determine the value of currency notes held in front of the camera, aiding users in financial transactions.
3. **Lucid Guidance and Assistance:** The proposed model serves as a lucid guide for visually impaired individuals, offering seamless navigation and functionality without causing any visual dilemma. Voice-based interaction ensures effortless access to various features, enhancing independence and confidence in daily activities.
4. **Road Assist Mode Activation:** Commanding “Start road assist mode” triggers the activation of the road assist feature, enabling real-time detection of vehicles, pedestrians, and hazards to enhance road safety for the user.
5. **SOS Feature:** In addition to the primary functionalities, the proposed application incorporates an SOS feature activated by voice command. When prompted, the application swiftly sends the user’s current location with accurate coordinates, along with their personal details, to the nearest hospital and police station, ensuring prompt assistance in emergency situations.

## 4 Results and Discussion

The training, validation, and testing are carried out on the Google Colab platform using the custom-labeled dataset. The following losses were considered for better understanding of the proposed quality of the training being done on the proposed model. The k-fold cross-validation technique was applied for training, validation, and testing for the proposed model.

**Box Loss:** It is used in object detection tasks, particularly in frameworks like YOLO. The box loss is responsible for penalizing inaccurate predictions of bounding box coordinates. It measures the discrepancy between the predicted bounding box coordinates and the ground truth bounding box coordinates for each object in the training data.

**Class Loss:** In the context of machine learning, particularly in tasks such as classification, “class loss” typically refers to the loss function used to measure the discrepancy between predicted class probabilities and the true class labels of the data.

**Object loss:** It is again used in object detection tasks, particularly in frameworks like YOLO. In object detection tasks, the model not only predicts the classes of objects present in an image but also their locations via bounding boxes. The object loss component of the loss function is responsible for penalizing inaccurate predictions of object presence (i.e., whether an object is present in a certain region of the image) and bounding box coordinates (i.e., the coordinates of the predicted bounding boxes). Minimizing this object loss during training helps ensure that the object detection model accurately predicts both the presence and location of objects within images.

## 4.1 Training Parameters

Images utilized for training and validation were stored in Roboflow, a platform designed for managing computer vision datasets. A train-validation-test split ratio of 75, 15, 10% was employed for the proposed model. Additionally, a batch size of 640 was used. The training process involved 200 epochs, allowing the model to iteratively learn and adjust its parameters over multiple passes through the dataset.

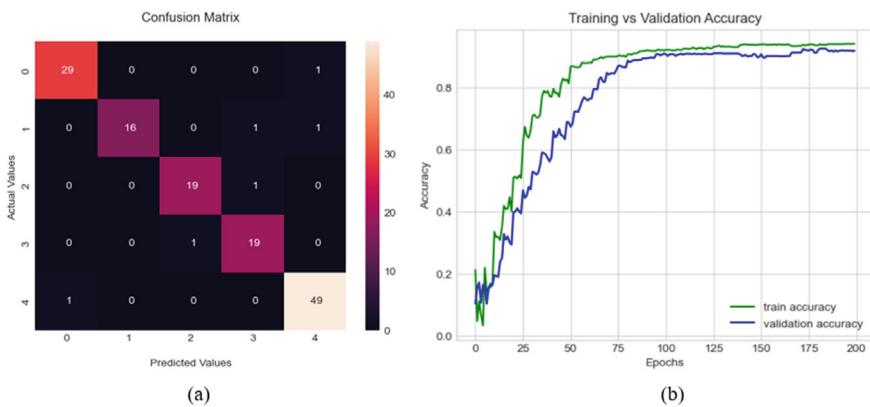
## 4.2 Validation Technique

To evaluate the performance of the proposed model, cross-validation techniques were employed, ensuring robustness and generalization capability. Additionally, standard evaluation metrics such as precision, recall, and F1-score were utilized to quantify the model’s performance across different use cases.

Overall, the proposed methodology encompasses a comprehensive approach to model selection, training, dataset management, and validation, aimed at deploying effective computer vision solutions to assist visually impaired individuals in various aspects of daily life. The results of the classification task are shown in Table 3. It shows the evaluation score on the test and validation dataset for different Image categories. Figure 4a and b shows the confusion matrix and training versus validation accuracy respectively. In the confusion matrix, the true positive box shows proposed model produces a good amount of correct labels towards the expected class. The training

**Table 3** Evaluation score on different image types

Image type	Precision	Recall	F1-Score	Average confidence score
Family and friends	0.95	0.95	0.95	0.78
Currencies	0.91	1	0.95	0.82
Bikes	0.93	0.93	0.93	0.81
Cycles	1	0.93	0.96	0.77
Roadside individuals	1	0.96	0.98	0.81

**Fig. 4** **a** Confusion matrix of classification. **b** Training versus validation accuracy plot

accuracy and validation accuracy provided in Fig. 4b revealed that the quality training has been provided to the proposed model.

In order to highlight the advantages of the proposed approach over existing methods, a comparative analysis is conducted, focusing on key parameters. These parameters include the model considered, the datasets utilized for implementation, and whether audio feedback methods were incorporated. The findings of this comparative analysis are summarized in the accompanying Table 4.

## 5 Conclusion and Future Scope

In this paper, we introduce an assistive technology designed to aid the visually impaired. To achieve this goal, we begin by creating a custom dataset encompassing a diverse range of objects and activities. The dataset undergoes preprocessing, involving resizing and background removal through image segmentation. Subsequently, each image is annotated with custom labels to facilitate training. We then employ a

**Table 4** Comparison of proposed approach with existing approaches

Paper	Objective	Dataset used	Results
[3]	Object detection for visually impaired	Custom dataset	Average accuracy of 96%
[5]	Single-stage object detection	PASCAL VOC, MS COCO	Multi-object classification with average accuracy 90%
[9]	Object detection, speech generation	Custom dataset, MS COCO	Acceptable object detection and speech generation
[10]	Object recognition for the visually impaired	Custom dataset	The accuracy in object recognition is 98.43%.
Proposed model	Object detection for visually impaired	Custom dataset	Average classification accuracy of 95.4 on fivefold cross-validation data%

pre-trained YOLO model for the training process. Additionally, voice-based accessibility features are integrated to enhance usability for visually impaired users. Our proposed model demonstrates satisfactory performance in correctly labeling images according to their expected categories. Furthermore, the results indicate minimal overfitting, thus affirming the quality of the training process.

To deploy the proposed model, the device required is a smartphone equipped with a camera and internet connectivity. The model will be a mobile application, with processing primarily handled by cloud servers to minimize the device load and utilize low battery consumption.

The future scope could incorporate features such as real-time object recognition, particularly for moving objects, and the ability to identify a safe distance while in motion. Additionally, enhancing the accuracy of object detection could be achieved by integrating alternative deep learning approaches.

## References

1. Honavar SG (2023) Indian Journal of Ophthalmology, Annual Report, 2022–23
2. Sharma T, Apoorva JHM, Lakshmanan R, Gogia P, Kondapaka M (2016) Navi: navigation aid for the visually impaired. In: 2016 international conference on computing, communication and automation (ICCCA). IEEE, pp 971–976
3. Mahesh TY, Parvathy SS, Thomas S, Thomas SR, Sebastian T (2021) Cicerone-a real time object detection for visually impaired people. In: IOP conference series: materials science and engineering, vol 1085. IOP Publishing, pp 012006
4. Dakopoulos D (2009) Tyflos: a wearable navigation prototype for blind & visually impaired; design, modelling and experimental results
5. Diwan T, Anirudh G, Tembhurne JV (2023) Object detection using yolo: challenges, architectural successors, datasets and applications. Multimedia Tools Appl 82(6):9243–9275

6. Kaur R, Singh S (2023) A comprehensive review of object detection with deep learning. *Digit Signal Process* 132:103812
7. Ying Yu, Wang C, Qiang F, Kou R, Huang F, Yang B, Yang T, Gao M (2023) Techniques and challenges of image segmentation: a review. *Electronics* 12(5):1199
8. Alkalai M, Lawgali A (2020) Image-preprocessing and segmentation techniques for vehicle-plate recognition. In: 2020 IEEE 4th international conference on image processing, applications and systems (IPAS). IEEE, pp 40–45
9. Guravaiah K, Bhavadeesh YS, Shwejan P, Vardhan AH, Lavanya S (2023) Third eye: object recognition and speech generation for visually impaired. *Procedia Comput Sci* 218:1144–1155
10. Md Atikur Rahman and Muhammad Sheikh Sadi (2021) IoT enabled automated object recognition for the visually impaired. *Comput Methods Programs Biomed Update* 1:100015
11. Sagana C, Keerthika P, Manjula Devi R, Sangeetha M, Abhilash R, Dinesh Kumar M, Hariharasudhan M (2021) Object recognition system for visually impaired people. In: 2021 IEEE international conference on distributed computing, VLSI, electrical circuits and robotics (DISCOVER). IEEE, pp 318–321
12. Hoeser T, Kuenzer C (2020) Object detection and image segmentation with deep learning on earth observation data: a review-part I: evolution and recent trends. *Remote Sens* 12(10):1667
13. Hoeser T, Bachofer F, Kuenzer C (2020) Object detection and image segmentation with deep learning on earth observation data: a review-part II: Applications. *Remote Sens* 12(18):3053
14. Torralba A, Russell BC, Yuen J (2010) Labelme: online image annotation and applications. *Proc IEEE* 98(8):1467–1484
15. Jocher G, Stoken A, Chaurasia A, Borovec J, Kwon Y, Michael K, Changyu L, Fang J, Skalski P, Hogan A et al (2021) Ultralytics/Yolov5: v6. 0-yolov5n’nano’models, roboflow integration, tensorflow export, OpenCV DNN support. Zenodo
16. Choi J, Gill H, Ou S, Song Y, Lee J (2018) Design of voice to text conversion and management program based on google cloud speech API. In: 2018 international conference on computational science and computational intelligence (CSCI). IEEE, pp 1452–1453
17. Adorf J (2013) Web speech API. KTH Royal Inst Technol 1:2013
18. Mahendru M, Dubey SK (2021) Real time object detection with audio feedback using yolo vs. yolo\_v3. In: 2021 11th international conference on cloud computing, data science & engineering (confluence). IEEE, pp 734–740
19. 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
20. Gaber M, Pastor GC, Omer A (2020) Speech-to-text technology as a documentation tool for interpreters: a new approach to compiling an ad hoc corpus and extracting terminology from video-recorded speeches. *TRANS: revista de traductología* (24):263–281

# Predicting COVID-19 Cases in India Using ARIMA, Prophet, LSTM and Data Analysis Using Power BI



Abhrajit Das

**Abstract** India confronted a serious challenge in the initial throes of the COVID-19 pandemic. This study addressed this by leveraging machine learning to analyse confirmed cases and support better decision-making. We compared the efficacy of Auto-Regressive Integrated Moving Average (ARIMA), Facebook Prophet and Long Short-Term Memory (LSTM) models on a Kaggle dataset, visualizing the results with Power BI to forecast future trends. The LSTM model, evaluated using Mean Absolute Percentage Error (MAPE), demonstrated superior accuracy in predicting case numbers. This research underlines the importance of data analysis during public health emergencies. Machine learning offers valuable insights for policymakers, empowering them to control outbreaks and allocate resources effectively. Ultimately, this paves the way for strengthened public health responses and preparedness for future pandemics.

**Keywords** Machine learning · Time series forecasting · Data science · Public health · Data visualization · Epidemiology

## 1 Introduction

The first instance of Novel coronavirus, which is also known as the Wuhan Virus or COVID-19, was reported in the middle of December 2019. The human-to-human transmission of nCov or the Coronavirus raised infected cases exponentially in this early stage. The World Health Organization (WHO) issued a worldwide health emergency on 30 January 2020 because of COVID-19 [1]. Morbidity and mortality rates for COVID-19 infection are unknown at an advanced stage, particularly for young and old people [2]. To control the widespread of COVID-19, government authorities took preventative actions and enforced curfews or shut down infested cities in most of the world. This helps the public authorities to implement social distancing

---

A. Das (✉)

Seidenberg School of Computer Science and Information Systems, Pace University, New York, USA

e-mail: [abhrajit.das@pace.edu](mailto:abhrajit.das@pace.edu)

among the people to prevent the spread of this novel virus [3]. The emergence of the COVID-19 pandemic in late 2019 has swiftly evolved into a global health crisis, profoundly impacting populations, economies and healthcare systems worldwide. Among the countries severely affected, India stands out due to its vast population and diverse socio-economic landscape [1–4]. From March to August 2020, India experienced a significant surge in COVID-19 cases, prompting stringent measures and public health interventions across various states to curb the spread of the virus [2–4]. This paper delves into the trajectory of COVID-19 in India during this critical period, focusing on data-driven analysis and predictive modelling using advanced machine learning techniques [5]. The study employs comprehensive data sourced from Kaggle [4], covering essential metrics such as confirmed cases, deaths and recoveries. These metrics were further analysed using Power BI [6–9], a powerful data visualization tool that enabled detailed insights into the pandemic's progression. The main contributions of the proposed work can be summarized as follows:

1. The research collected a comprehensive COVID-19 dataset from Kaggle (March to August 2020), serving as a robust foundation for time series analysis [10–12], detailed data visualization using Power BI and model training using Python.
2. The study evaluates and compares ARIMA, Prophet and LSTM models using Mean Absolute Percentage Error (MAPE) as the primary performance metric, demonstrating their applicability in real-world public health scenarios.
3. The proposed scheme has the potential to provide a predictive tool for assessing the status of COVID-19 infection and enable government and health workers to make better decisions to reduce mortality.

The structure of this paper is organized as follows: Sect. 2 reviews related work and theoretical background. Section 3 describes the data and methodology used for analysis and predictive modelling. Section 4 presents the results of the analysis, compares the performance of the predictive models and discusses the implications of the findings for public health policy and future research. Finally, Sect. 5 concludes the paper and outlines potential directions for future work.

## 2 Related Works

While machine learning holds promise for predicting infectious disease outbreaks [13–22], existing research on COVID-19 forecasting in India has limitations. Many studies lack comprehensive comparisons of different models, particularly within the specific socio-economic and demographic context of India [5]. Additionally, existing research often does not consider how well models adapt to rapidly evolving data patterns during a pandemic. This study addresses these gaps by comparing the performance of three prominent machine learning models: ARIMA, Prophet and LSTM. We evaluate their accuracy in forecasting COVID-19 cases in India and assess their ability to adapt to changing data trends. Furthermore, we acknowledge the

importance of data visualization tools like Power BI in interpreting model predictions and informing public health decisions.

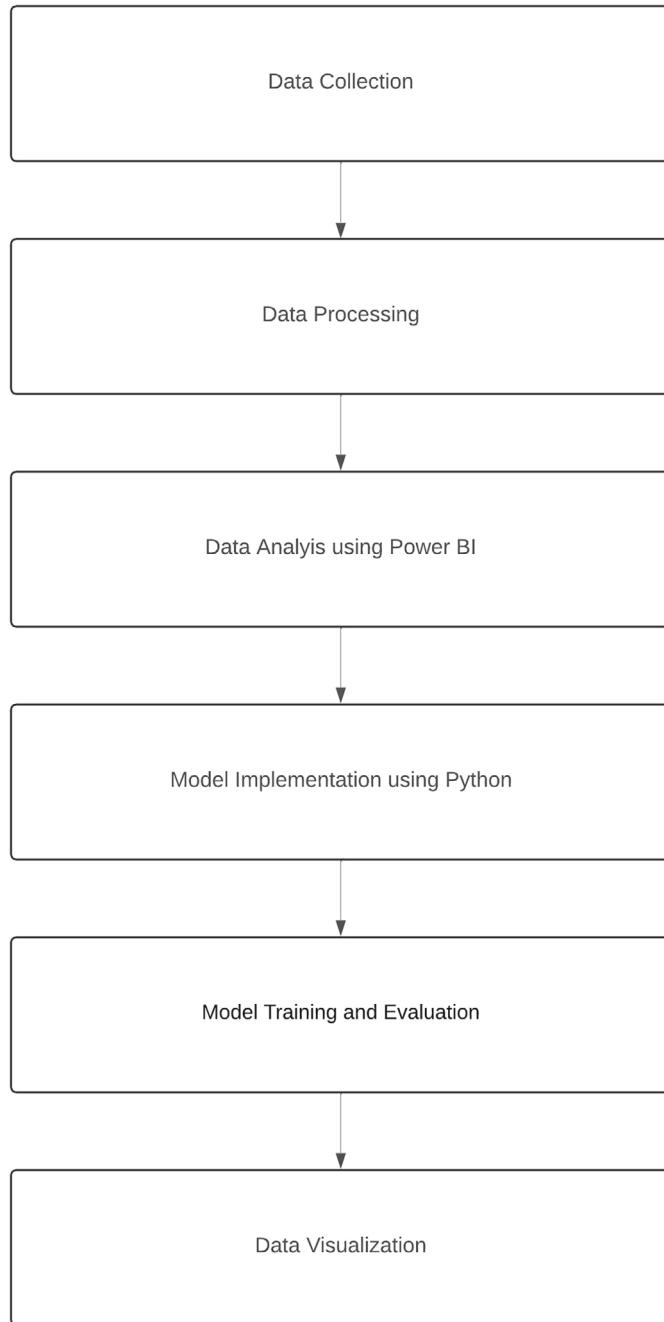
### 3 Proposed Scheme

The proposed scheme involves the implementation and comparison of three advanced machine learning models—ARIMA [15–17], Prophet [18–20] and LSTM [21, 22] to predict the number of COVID-19 cases in India. Each model was meticulously trained and evaluated based on its accuracy in forecasting actual case counts. The primary performance metric used for this evaluation was the Mean Absolute Percentage Error (MAPE), which provides a clear measure of each model's predictive accuracy [23].

By comparing the performance of these models using MAPE, the study aimed to identify the most reliable and accurate predictive model. The evaluation process not only focused on the numerical accuracy of predictions but also considered the models' ability to adapt to evolving data patterns over time. This comprehensive analysis provided insights into the strengths and limitations of each model, contributing to a better understanding of their applicability in real-world public health scenarios. The results of this comparative analysis are intended to inform health consultants and policymakers about the most effective modelling techniques for predicting COVID-19 trends (Fig. 1).

#### 3.1 *The ARIMA (Auto-Regressive Integrated Moving Average) Model*

The ARIMA (Auto-Regressive Integrated Moving Average) Model is a classic time series model known for its simplicity and effectiveness in capturing linear relationships in data. ARIMA, which stands for Auto-Regressive Integrated Moving Average, is defined by three parameters: p, d and q. The parameter p (Auto-Regressive part) signifies the number of lag observations included in the model, essentially using past values to predict the current value [15]. The parameter d (Integrated part) indicates the number of times the raw observations are differenced to make the time series stationary, thus stabilizing the mean by eliminating trends and seasonality [16]. The parameter q (Moving Average part) represents the size of the moving average window, which smooths out the noise in the data by averaging past forecast errors. ARIMA models are particularly effective in scenarios where data exhibits linear trends and correlations over time, making them powerful tools for predicting short-term trends in various fields, including economics, finance and epidemiology. Despite their simplicity, ARIMA models can provide robust forecasts when appropriately parameterized, offering valuable insights by decomposing time series into understandable components [17].



**Fig. 1** Processing steps in the proposed methodology

The equation for the model is as follows:

$$Y_t = c + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (1)$$

where

$Y_t$  is the time series value at time  $t$ ,

$c$  is a constant,

$\Phi$  and  $\theta$  are parameters,

$\epsilon$  is white noise.

### 3.2 *Prophet Model*

Prophet Model a time series model by Facebook, is adept at handling seasonality and holiday effects in time series data. Prophet is a decomposable time series model with three main components: trend, seasonality and holidays [18]. The trend component captures the overall direction of the data over time, whether it is increasing, decreasing or remaining stable. The seasonality component accounts for periodic patterns that repeat at regular intervals, such as daily, weekly or yearly cycles. The holidays component incorporates the effects of holidays and special events that can cause significant deviations in the data. Prophet's strength lies in its ability to fit complex data patterns and make accurate forecasts by combining these components, making it particularly useful for scenarios where data is influenced by seasonal variations and specific events [19].

The equation for the model is as follows:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (2)$$

where

$g(t)$  is a piecewise linear or logistic growth curve for modelling non-periodic changes in time series.

$s(t)$  is periodic changes (e.g. weekly/yearly seasonality).

$h(t)$  is the effects of holidays that occur on potentially irregular schedules over one or more days [20].

### 3.3 The LSTM (Long-Short Term Memory) Model

The LSTM (Long-Short Term Memory) Model a type of recurrent neural network (RNN), stands out for its ability in sequence prediction tasks. This strength stems from its core component, the memory cell. Unlike traditional RNNs that struggle to remember information from earlier in a sequence, the LSTM's memory cell persists over time. This allows the model to retain crucial details and use them for future predictions [21]. This capability makes LSTMs particularly valuable for tasks where context and the order of information are essential, such as natural language processing, time series forecasting and speech recognition. Additionally, LSTMs excel at overcoming the vanishing gradient problem that plagues traditional RNNs. This enables them to effectively learn from and make predictions based on lengthy data sequences. The combination of its memory capabilities and ability to handle long-term dependencies makes LSTMs a powerful tool for various applications, especially those involving complex patterns within data sequences [22].

The equations for the model are as follows:

$$I_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (3)$$

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (4)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (5)$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_C) \quad (6)$$

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \quad (7)$$

$$h_t = o_t \tanh(C_t) \quad (8)$$

where

$i_t$ ,  $f_t$  and  $o_t$  are the input, forget and output gates respectively,

$\tilde{C}_t$  is the cell state,

$C_t$  is the updated cell state,

$h_t$  is the hidden state [21].

### 3.4 MAPE (Mean Absolute Percentage Error)

MAPE (Mean Absolute Percentage Error) is a widely used metric for measuring the accuracy of a forecast or prediction model. It is particularly valued for its simplicity and interpretability, providing a percentage error that indicates the average deviation of the predicted values from the actual values. This makes it easier to understand and communicate the performance of a model in practical terms. MAPE is calculated as the average of the absolute percentage errors between the actual and predicted values. It quantifies the accuracy of a model as a percentage, giving an intuitive sense of how far off the predictions are from the actual values on average [23].

The formulae for MAPE is

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \times 100 \quad (9)$$

where

$n$  is the number of observations.

$A_t$  represents the actual value at time  $t$ .

$F_t$  represents the forecasted (predicted) value at time  $t$ .

$\left| \frac{A_t - F_t}{A_t} \right|$  denotes the absolute value.

MAPE = 0% indicates Perfect accuracy, the predicted values exactly match the actual values.

Lower MAPE indicates better model accuracy, as the average percentage error is smaller.

Higher MAPE indicates poorer model accuracy, as the average percentage error is larger.

## 4 Performance Analysis

This section represents the performance of the proposed work.

### 4.1 Experimental Setup

The study utilized the following configuration for model development and analysis:

#### Hardware:

- RAM: 16 GB

- Processor: 11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40 GHz
- Operating System: Windows 11

### **Software:**

- Python 3.8
- Jupyter Notebook for model development and execution
- Power BI [6] for Data Visualization

The experiment involved preprocessing the data to prepare it for modelling. This likely included cleaning, handling missing values and potentially feature engineering. Following preprocessing, the chosen machine learning models were implemented in Python within the Jupyter Notebook environment.

## **4.2 Dataset Overview**

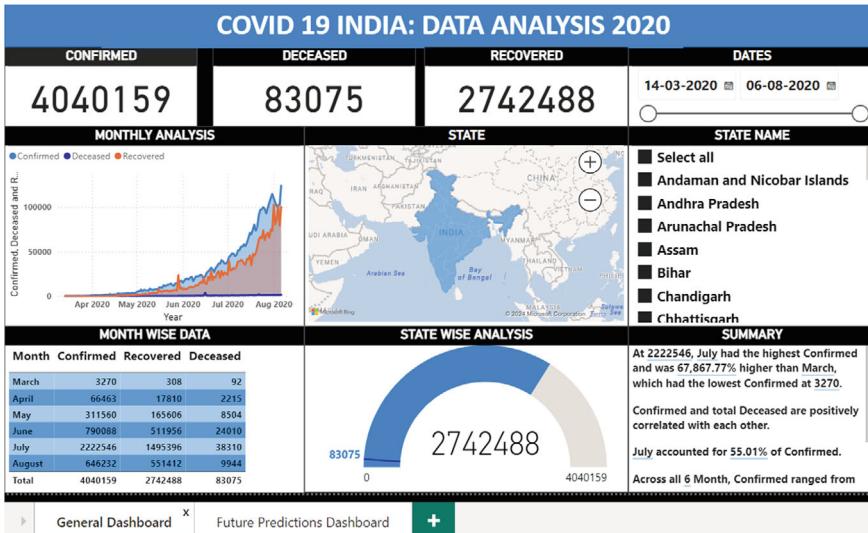
This study utilizes a COVID-19 dataset encompassing daily cases, deaths and recoveries across 28 Indian states and 8 union territories (March–August 2020). With 5,694 entries, it offers granular-level insights for analysis and forecasting models, crucial for understanding the pandemic's progression and evaluating prediction models' effectiveness [4].

## **4.3 Results and Discussions**

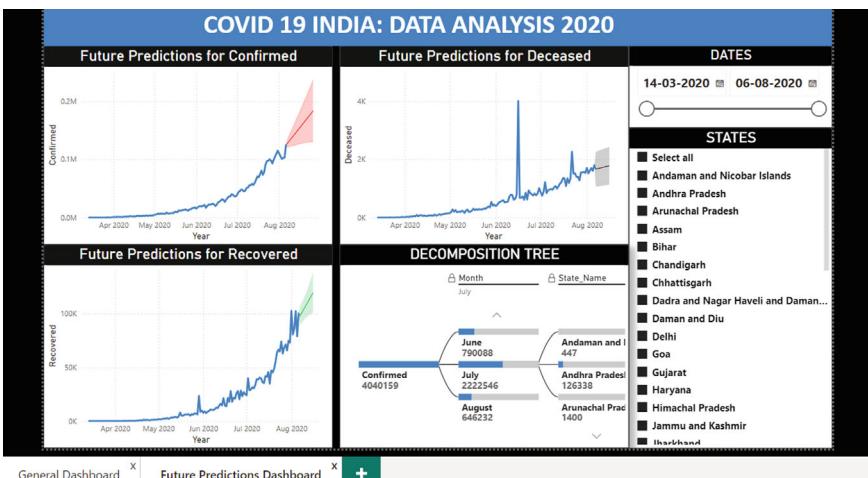
This study analysed COVID-19 data in India from March to August 2020, employing various techniques for exploration, visualization and prediction. A Power BI dashboard (Fig. 2) served as the central hub for data exploration, offering functionalities like time series analysis, state-wise breakdowns and future predictions (Fig. 3). This interactive platform facilitated understanding trends, hotspots, geographical spread and the impact of interventions.

The analysis revealed significant regional variations in COVID-19 cases. Maharashtra emerged as the most affected state (479,779 cases), likely due to factors like population density and healthcare disparities [1–4]. Conversely, states with lower population densities or proactive containment measures generally reported fewer cases. Peak periods highlighted critical moments of heightened transmission, informing resource allocation for future outbreaks.

Three machine learning models (ARIMA, Prophet, LSTM) were employed to forecast confirmed COVID-19 cases from July 13 to August 6, 2020 (Table 1). The LSTM model achieved the highest accuracy (MAPE: 0.0617) due to its ability to capture complex patterns (Fig. 6) compared to ARIMA (MAPE: 0.375, Fig. 4) and Prophet (MAPE: 0.228, Fig. 5). While Prophet effectively handled seasonality (advantageous for COVID-19 forecasting) [18–20], LSTM's strength



**Fig. 2** COVID-19 interactive dashboard on power BI showcasing time series analysis, geographical heat maps, comparative analysis and trend analysis



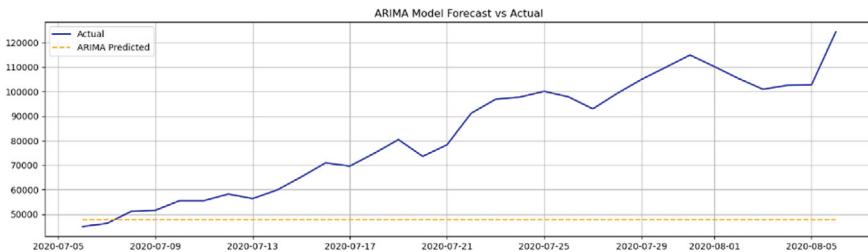
**Fig. 3** Dashboard for future predictions of COVID-19 cases in India: confirmed, deceased and recovered

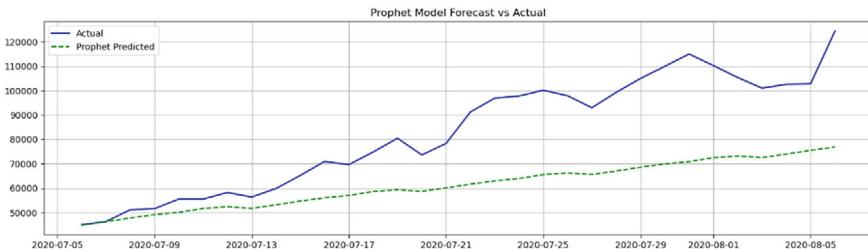
lies in capturing long-term dependencies and non-linearities, crucial for accurate predictions [21, 22].

Overall, this study highlights the potential of Power BI for data exploration and visualization, and machine learning, particularly LSTM, for forecasting COVID-19 cases. These tools can be valuable assets for public health officials in understanding

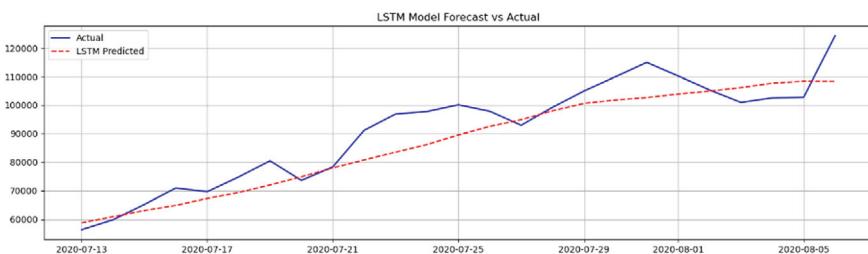
**Table 1** Comparison of actual and predicted values for ARIMA, prophet and LSTM models

Date	Actual data	ARIMA predicted	Prophet predicted	LSTM predicted
2020-07-13	56,356	47,795	51,738	58,761
2020-07-14	59,834	47,795	53,145	60,911
2020-07-15	65,214	47,795	54,707	63,083
2020-07-16	70,936	47,795	56,046	64,875
2020-07-17	69,640	47,795	57,016	67,295
2020-07-18	74,822	47,795	58,619	69,363
2020-07-19	80,470	47,795	59,306	72,009
2020-07-20	73,612	47,795	58,763	74,870
2020-07-21	78,340	47,795	60,080	78,010
2020-07-22	91,202	47,795	61,641	80,784
2020-07-23	96,886	47,795	62,981	83,525
2020-07-24	97,776	47,795	63,950	86,164
2020-07-25	100,144	47,795	65,553	89,546
2020-07-26	97,864	47,795	66,241	92,565
2020-07-27	92,968	47,795	65,607	94,900
2020-07-28	99,262	47,795	67,014	98,031
2020-07-29	104,958	47,795	68,576	100,625
2020-07-30	109,936	47,795	69,915	101,805
2020-07-31	114,972	47,795	70,885	102,666
2020-08-01	110,234	47,795	72,488	103,904
2020-08-02	105,344	47,795	73,175	104,937
2020-08-03	100,976	47,795	72,542	106,148
2020-08-04	102,564	47,795	73,949	107,699
2020-08-05	102,774	47,795	75,510	108,378
2020-08-06	124,340	47,795	76,850	108,289

**Fig. 4** Comparison of actual versus predicted daily confirmed COVID-19 cases in India using ARIMA model



**Fig. 5** Comparison of actual versus predicted daily confirmed COVID-19 cases in India using prophet model



**Fig. 6** Comparison of actual versus predicted daily confirmed COVID-19 cases in India using LSTM model

the pandemic's dynamics, planning interventions and allocating resources effectively. Continuous monitoring and adaptation of strategies remain essential, as unforeseen circumstances can influence forecasts [5].

#### 4.4 Exploratory Data Analysis (EDA) and Data Visualization with Power BI

A Power BI dashboard effectively visualized COVID-19 data (Fig. 2), offering an interactive interface for exploring trends, hotspots and geographical spread over time. Users could filter data by specific timeframes and locations, providing a granular view of the pandemic's progression. Highlighting key metrics and trends, the dashboard aided in understanding the impact of interventions and pinpointing periods of rapid case increases. This visualization tool proved essential for health officials and researchers, providing actionable insights to guide public health strategies and resource allocation. It also fostered public awareness by disseminating critical information. Furthermore, the dashboard offered functionalities for future prediction (Fig. 3), allowing users to forecast the number of confirmed, recovered and deceased cases. Additionally, a decomposition tree facilitated detailed analysis by breaking down total confirmed cases by month and state name. Overall, the Power

BI dashboard enhanced data transparency and data-driven decision-making during the pandemic. Key features of the Power BI dashboards included:

**Monthly Analysis:** Visualizing the daily trends in confirmed cases, deaths and recoveries to understand the progression of the pandemic.

**State-wise Analysis:** Identifying peaks and troughs in case numbers to pinpoint critical transmission periods.

**Detailed Summary:** Provided detailed summary of the entire dashboard.

**Future Predictions:** Forecasting the number of confirmed, recovered and deceased cases for the future (Fig. 3).

**Decomposition Tree:** Breaks down total confirmed cases by month and state name for detailed analysis.

## 4.5 COVID-19 Impact and Statistical Summary

During the period from March to August 2020, India reported a total of 4,040,159 confirmed COVID-19 cases, with 83,075 deaths and 2,742,488 recoveries. Maharashtra emerged as the most severely affected state, recording 479,779 confirmed cases, 16,791 deaths and 320,893 recoveries, followed by Tamil Nadu with 273,969 confirmed cases, 4,456 deaths and 215,056 recoveries. The analysis reveals variations in COVID-19 impact across states in India. States like Maharashtra, Tamil Nadu and Delhi reported the highest total confirmed cases, reflecting the regional disparities in healthcare infrastructure and population density [1–4]. On the other hand, states with lower population densities or proactive containment measures generally recorded fewer cases and deaths. The peak values highlight critical periods of transmission and healthcare strain, guiding resource allocation and policy interventions. The future predictions in Fig. 3 suggested that there would be a significant rise in the number of confirmed and recovered cases with a minimal increase in deceased cases over time, but limitations exist as unforeseen circumstances can influence forecasts. Continuous monitoring and adaptation of public health strategies remain paramount in mitigating the pandemic's impact [3].

## 4.6 Machine Learning Models

Table 1 represents a detailed comparison of the actual versus predicted values for confirmed COVID-19 cases in India. This comparison spans the period from July 13, 2020 to August 6, 2020, and utilizes three different predictive models: ARIMA, Prophet and LSTM. Each row is dedicated to one of these models, showcasing the effectiveness and accuracy of the predictions during this specified timeframe.

Figures 4, 5 and 6 illustrate the actual versus predicted trends of confirmed COVID-19 cases in India from July 13, 2020 to August 6, 2020. These figures depict line graphs where the x-axis represents the dates within this range, and the y-axis indicates the number of confirmed COVID-19 cases. Figures 4, 5 and 6 all depict predicted values (orange in Fig. 4, green in Fig. 5 and red in Fig. 6) compared to the actual values (blue in all figures).

#### 4.7 ARIMA Model

ARIMA Model demonstrated moderate performance in predicting the number of confirmed COVID-19 cases in India. It was effective in capturing linear trends and short-term variations within the dataset. However, the model encountered difficulties when dealing with non-linear patterns and seasonal fluctuations, which affected its accuracy in making long-term forecasts. Despite these challenges, the ARIMA model achieved a Mean Absolute Percentage Error (MAPE) of 0.375. This indicates that while the model was somewhat effective in capturing general trends, it fell short when it came to making precise predictions. Figure 4 and Table 1 detail the ARIMA model's performance metrics and insights, offering a comprehensive view of its strengths and weaknesses in prediction [15–17].

#### 4.8 Prophet Model

Prophet Model exhibited robust performance in forecasting the number of confirmed COVID-19 cases in India. One of its key strengths is its ability to effectively handle seasonality and holiday effects, which are critical factors in accurately modelling the spread of the virus. The model achieved a Mean Absolute Percentage Error (MAPE) of 0.228, 15 indicating a significant improvement over the ARIMA model. Prophet excelled in capturing both short-term fluctuations and long-term trends, demonstrating its flexibility in modelling complex data patterns. This flexibility contributed to its superior accuracy in predicting the dynamics of COVID-19, making it a reliable tool for understanding and forecasting the pandemic's progression. Figure 5 and Table 1 together illuminate the Prophet model's performance and insights, highlighting its effectiveness relative to other models [18–20].

#### 4.9 LSTM Model

LSTM Model emerged as the most accurate model for forecasting the number of confirmed COVID-19 cases in India. Leveraging its advanced capacity to capture non-linear dependencies and long-term trends, LSTM demonstrated exceptional

performance in predicting the dynamics of the pandemic. The model achieved a Mean Absolute Percentage Error (MAPE) of 0.0617, surpassing both the ARIMA and Prophet models in terms of accuracy. This low MAPE indicates the LSTM model's high precision in forecasting. Its ability to learn from historical data and adapt to evolving trends was instrumental in accurately predicting COVID-19 case numbers across different states. Figure 6 and Table 1 showcase the LSTM model's superior predictive capabilities [21, 22].

## 5 Conclusions and Future Work

This study explored India's initial COVID-19 wave (March-August 2020). Power BI empowered researchers with data exploration and visualization, uncovering regional variations in cases linked to population density and healthcare disparities. Power BI's forecasting functionalities were also explored, providing initial predictions for future trends. Further analysis employed machine learning models to refine case forecasts. The LSTM model emerged as the most accurate (MAPE: 0.0617), surpassing ARIMA and Prophet due to its ability to handle complex patterns. The LSTM model's accuracy suggests its potential to inform public health strategies and resource allocation during future outbreaks.

Future research could refine models with additional data (demographic, socio-economic) and explore ensemble techniques for enhanced accuracy. Real-time data integration and scenario-based simulations could provide valuable insights for pandemic response strategies.

## References

1. World Health Organization (WHO) (2020) Coronavirus disease (COVID-19) pandemic. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
2. Eissa D, Rashed E, Eissa M (2022) A Study of morbidity and mortality from COVID-19 in India. *SciMedicine J* 4:25–38. <https://doi.org/10.28991/SciMedJ-2022-0401-03>
3. Makandar S, Prabhakar S (2021) COVID-19 lockdown in India and its impact on people's livelihoods. *PalArch's J Archaeol Egypt/Egyptol* 18:381–388
4. Imdevskp (2020) COVID-19 India state-wise dataset for confirmed, deceased, and recovered from March 2020 to August 2020. [https://www.kaggle.com/datasets/imdevskp/covid19-corona-virus-india-dataset?select=state\\_level\\_daily.csv](https://www.kaggle.com/datasets/imdevskp/covid19-corona-virus-india-dataset?select=state_level_daily.csv)
5. Sah S, Surendiran B, Dhanalakshmi R, Mohanty S, Alenezi F, Polat K (2022) Forecasting COVID-19 pandemic using prophet, ARIMA, and hybrid stacked LSTM-GRU models in India. *Comput Math Methods Med* 2022:1–19. <https://doi.org/10.1155/2022/1556025>
6. Microsoft Corporation (2020) Power BI documentation. <https://docs.microsoft.com/en-us/power-bi/>
7. Abinaya S (2024) Data visualization using power BI. *Int Sci J Eng Manag* 3:1–9. <https://doi.org/10.55041/ISJEM01536>
8. Aravinthan B (2024) Property future price estimation using ML, power BI time series analysis and forecasting. *Int J Sci Res (IJSR)* 13:619–622. <https://doi.org/10.21275/SR24330134715>

9. Rajput S (2024) Comparison between tableau and power BI: a case study of COVID-19. *Int J Sci Res Eng Manag* 8:1–5. <https://doi.org/10.5504/IJSREM32733>
10. Namasudra S, Dhamodharavadhani S, Rathipriya RG, Crespo RG, Moparthi NR (2023) Enhanced neural network-based univariate time series forecasting model for big data. *Big Data*. <https://doi.org/10.1089/big.2022.0155>
11. Box GEP, Jenkins GM, Reinsel GC (1994) Time series analysis: forecasting and control. John Wiley & Sons, New York
12. Makridakis S, Wheelwright SC, Hyndman RJ (1998) Forecasting: methods and applications, 3rd edn. John Wiley & Sons, New York
13. Dua D, Graff C (2019) UCI machine learning repository. School of Information and Computer Sciences, University of California, Irvine. <http://archive.ics.uci.edu/ml>
14. Pedregosa F et al (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
15. Singh S, Parmar K, Kumar J, Kaur J (2022) Prediction of confirmed, recovered and casualties' cases of COVID-19 in India by autoregressive integrated moving average (ARIMA) models. In: Proceedings of the international conference on intelligent systems and signal processing. Springer, New York, pp 305–312 (2022). [https://doi.org/10.1007/978-3-030-72834-2\\_6](https://doi.org/10.1007/978-3-030-72834-2_6)
16. Alzahrani S, Saadeh R, Abdoon M, Qazza A, Guma F, Berir M (2024) Numerical simulation of an influenza epidemic: prediction with fractional SEIR and the ARIMA model. *Appl Math & Inf Sci* 18:1–12. <https://doi.org/10.18576/amis/180101>
17. Namasudra S, Dhamodharavadhani S, Rathipriya R (2021) Nonlinear neural network-based forecasting model for predicting COVID-19 cases. *Neural Process Lett*. <https://doi.org/10.1007/s11063-021-10495-w>
18. Kapse P, Timande P, Bramhankar A, Rewatkar S, Khandait S (2022) Analysis & prediction of COVID-19 using prophet model. *Int J Res Appl Sci Eng Technol* 10:1352–1356. <https://doi.org/10.22214/ijraset.2022.42481>
19. Tulshyan V, Sharma D, Mittal M (2020) An eye on the future of COVID'19: prediction of likely positive cases and fatality in India over a 30 days horizon using prophet model. *Disaster Med Public Health Prep* 16. <https://doi.org/10.1017/dmp.2020.444>
20. Kumar P, Sharma R, Singh S (2021) Predictive analysis of real-time strategy using Facebook's prophet model on COVID-19 dataset of India. *J Pharm Res Int* 33(51A):305–312. <https://doi.org/10.9734/jpri/2021/v33i51A33496>
21. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
22. Chandra R, Jain A, Chauhan D (2021) Deep learning via LSTM models for COVID-19 infection forecasting in India
23. Robeson S, Willmott C (2023) Decomposition of the mean absolute error (MAE) into systematic and unsystematic components. *PLoS ONE* 18:e0279774. <https://doi.org/10.1371/journal.pone.0279774>

# Classification of Multi-Labeled Retinal Diseases in Retinal Fundus Images Using CNN Model ResNet18



Kowju Gayatri and Birendra Biswal

**Abstract** Fundus images are frequently utilized by medical professionals such as ophthalmologists and are quite useful in detecting various retinal abnormalities. They utilized this to diagnose a variety of eye diseases, including cataracts, diabetic retinopathy, and glaucoma. Automatic diagnosis of retinal diseases is a highly challenging task today. Retinal fundus images play a vital role in providing valuable information for ophthalmologists to the fast and accurate diagnoses. Early and accurate diagnosis of retinal diseases is crucial for timely intervention and treatment. The majority of people suffer from a lack of accurate diagnosis to prevent their vision loss. This research work employs a ResNet18 model to classify these fundus images into multi-labeled categories. Fundus images consist of four categories such as cataracts, diabetic retinopathy, glaucoma, and normal cases and these are major causes of vision impairment worldwide. The dataset comprises 4217 retinal fundus images belonging to four classes collected from Kaggle. Our proposed ResNet18 model is trained on this dataset, and we split the dataset into train-test-validation parts. The model has been trained & validated using Kaggle datasets. Finally, the proposed model, achieved 100 & 94% accuracy on the training and validation dataset.

**Keywords** F1-score · Convolutional neural network(CNN) · Diabetic retinopathy-DR · Glaucoma

---

K. Gayatri   
Andhra University, Visakhapatnam, India  
e-mail: [gayatrisantosj@gvpce.ac.in](mailto:gayatrisantosj@gvpce.ac.in)

K. Gayatri · B. Biswal  
Centre for Medical Imaging Studies, Department of ECE, Gayatri Vidya Parishad College of Engineering (A), Visakhapatnam, India  
e-mail: [birendrabiswal@gvpce.ac.in](mailto:birendrabiswal@gvpce.ac.in)

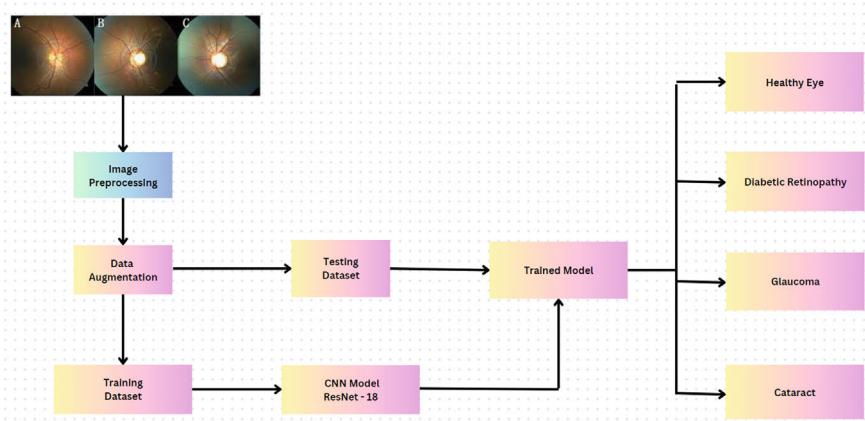
## 1 Introduction

A little compromise of quality in the human eye can have a negative impact on an individual's productivity and quality of life. Millions of people around the world suffer from retinal abnormalities, and if left untreated, they can lead to vision loss. Early detection and proper treatment can stop the progression of retinal diseases and may prevent many people from losing their vision altogether. Veteran ophthalmologists use retinal images taken with fundus cameras or OCTs to detect various retinal diseases. However, limited access to medical experts and infrastructure in developing countries, especially in rural areas, emphasizes the need for automated retinal disease detection. Glaucoma, DR, cataracts highly affect the surface of the retina which can significantly impact vision and overall health. Glaucoma is an irreversible disease that causes permanent vision loss, it can harm the optic nerve system, often due to elevated pressure within the eye. As reported by WHO, approximately 3% of the world's population aged 40 to 80 years old has glaucoma, which equates to about 80 million people worldwide. Diabetic retinopathy is associated with diabetes difficulty that will create problems in retina blood vessels, these can be impacted by sugar levels and potential vision loss. The International Diabetes Federation (IDF) reports that roughly 33% of the population with diabetes have some degree of diabetic retinopathy. As of 2021, around 537 million adults were affected with diabetes universally, indicating a substantial number of individuals at risk of developing this condition. When a cataract develops, the natural lens of the eye becomes clouded, impairing visual acuity and blurring vision. As per the National Eye Institute records, by 25–80 years old, above 50% of people suffer from the impact of cataracts. Worldwide, it is estimated that millions of people are affected by cataracts, with a significant portion of the global population experiencing this condition during their lifetime. Each of these conditions requires proper diagnosis, treatment, and ongoing management by healthcare professionals to maintain optimal eye health and preserve vision. Individuals need to undergo regular eye exams to detect these conditions early and seek appropriate care when necessary. The study aims to denote the capacity of CNN methods, particularly ResNet, in accurately diagnosing retinal diseases for timely intervention and treatment.

## 2 Literature Review

Excellent work has been done by researchers in the fields of Glaucoma, diabetic retinopathy, and cataracts they have proposed and implemented various advanced methods to categorize retinal diseases. Indra Weni et al. [1] proposed to determine the cataract disease identification using the optimal Convolutional Neural Networks. The data used consists of images of eyes, including normal & cataracts. They have concentrated on the LR and trained for more epochs and they achieved 95% accuracy but they have not worked on any other diseases. Supriya Mishra et al. [2] proposed a

Deep Learning model DenseNet on a dataset comprising 3662 images they classified the stages of DR and obtained an accuracy of 96% this work is limited to DR only. Mohammed Ghazal et al. [3] proposed CAD system to detect DR they tried to find DR optimal conditions it can detect only on OCT images. Sadikul Alim Toki et al. [4] concentrated on diagnosing the many forms of ailments. Their purpose was to determine whether or not the person's eye was harmed, but not to disclose the disease category. They achieved 95% accuracy using a simple CNN architecture. Syed Al E Hassan et al. [5] employed pre-trained deep CNN models on OCT images, focusing on Central Serous Retinopathy (CSR), and achieved an accuracy of 99.64% for AlexNet, ResNet18, and GoogleNet, indicating that it is more sophisticated. Veena et al. [6] proposed an architecture that combines two Convolutional models for Optic-cup and Optic-Disk separately and this model includes the identification of only one retinal disease (glaucoma). Yaqoob et al. [7] proposed a Random Forest classifier over ResNet50 for grading diabetic retinopathy. He achieved 96% accuracy on the Messidor-2 dataset, but only 76% on the EyePACS database with five classifications. Sushma K. Sattigeri et al. [8] suggested DL models for classifying cataracts, bulging eyes, crossed eyes, uveitis, and conjunctivitis. Training for single-eye and two-eye pictures is carried out using two distinct models. One model uses two-eye imagery to forecast disorders such as crossed and bulging eyes. The other model predicts illnesses such as cataracts and conjunctivitis/uveitis based on single-eye imagery. They achieved an accuracy of 96% for the dataset of single-eye images and 92.31% for the dataset of two-eye images. Whereas [9] employed an MLP classifier to detect diabetic retinopathy in eye photographs, any retinal image affected by another eye ailment, such as glaucoma or cataracts, could not be identified. They employed the MLP NN classifier and achieved 100 percent accuracy on the validation dataset. Turimerla Pratap [10] presented a computer-aided automatic method for obtaining precise information about cataract phases from fundus images. They utilized pre-trained CNN. The four-stage cataract accuracy achieved 92.91%. Jayesh Vasudeva et al. [11] provided a summary of the most recent techniques established to identify and segment the glaucoma in eye. Glaucoma is difficult to identify with a single test, hence ophthalmologists undertake numerous tests to identify a patient as glaucomatous or not whereas CDR is often used to assess optic nerve damage; a higher CDR suggests a greater risk of glaucoma. These methods are based on deep learning, and feature engineering techniques such as thresholding, and pixel categorization. Hamza Mustafa et al. [12] presented the deep CNN models ResNet50 and DenseNet-121, which reached 95.58 accuracy. An ML classifier was used to improve accuracy. As a result, they attained the accuracy of every state-of-the-art. Several deep CNN models were applied to diverse eye datasets. In this study, we discovered that the ResNet18 model produced good results for classifying retinal diseases. Lorick Jain et al. [13] proposed LCDNet architecture methodology The model was tested on a pair of datasets, incorporating actual patient fundus images of the retina acquired from a nearby hospital. As a result, they achieved an accuracy of 96.5%. In this research work the author gave a detailed analysis of eye conditions such as whether the disease is present or not but doesn't label the type of disease. Even though as discussed, deep learning techniques have achieved excellent results, still they

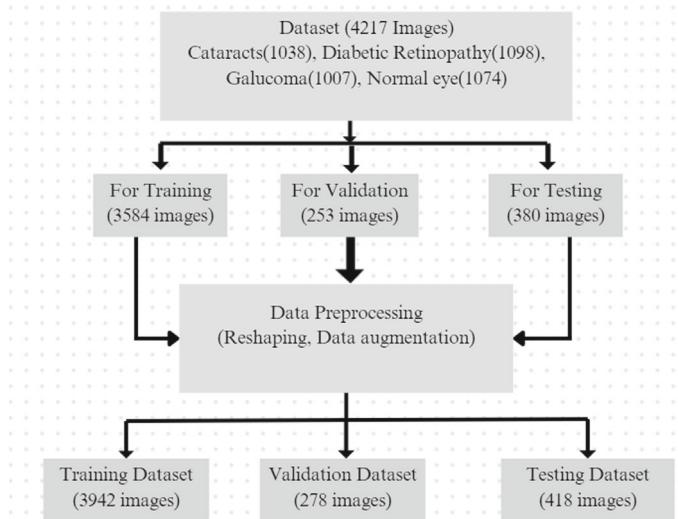


**Fig. 1** Methodology block diagram

suffer from many challenges because retinal diseases are classified separately. This research work investigates the various diseases in retinal fundus images. However, the proposed Resnet18 model shall classify the multiple retinal diseases simultaneously using a convolution neural network as the backbone. We have applied different CNN techniques on the same dataset, but the Resnet18 has given the best results with less computation complexity as well as it improved the quality of the parameters such as accuracy, Precision, Recall, and F1-score (Fig. 1).

## 2.1 Data Acquisition

Data acquisition involves the collection of unprocessed data from diverse sources for assessment, processing, and storing. This step is vital in the data life cycle it plays an essential part in various sectors, including scientific study, Business analytics, healthcare, and neural networks. Acquiring high-quality, diverse data is crucial for accurate and effective decision-making and evaluation and it includes capturing high-resolution retinal images using modalities such as Photography, OCT, and Fluorescein angiography to obtain detailed representations of retinal structures and abnormalities. Data acquisition involves labeling and annotating retinal images with ground truth information related to disease presence and severity, which is essential for training deep learning models for disease classification. In this stage we have collected the dataset through the Kaggle source with “<https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification>”.



**Fig. 2** Dataset structure

## 2.2 *Image Preprocessing*

Image preprocessing is a collection of approaches used on digital images before they are sent into a Deep learning or vision-based system for evaluation or processing. The preprocessing of images aims to improve image quality and extract useful information, and standard formats for machine learning models. Preprocessing can enhance the contrast and sharpness of retinal images, improving the visibility of key features such as blood vessels, lesions, and the optic disk, which are essential for disease classification. It focuses on enhancing the image grade as per requirements. These methods can regenerate extra images or produce features from the source image (Fig. 2).

## 2.3 *Data Augmentation*

Augmentation is an activity that makes alterations to an image, generated images can be treated as distinct ones by the system, whereas human beings might be considered like same. Data augmentation increases the size of the training dataset by applying various domain-specific transformations to the input data. The transformations are rotating the image by a certain angle, flipping the image, zooming the images, Adjusting RGB values, transforming to grayscale, or applying other color variations. Data augmentation prevents overfitting by expanding and diversifying the training dataset. This reduces the likelihood of the model remembering the data.

and being unable to generalize to new samples. As a whole, data augmentation is effective. This tool enhances the quality and reliability of machine learning models in various fields and applications. In our model after data augmentation the final dataset size expanded to 4638 fundus images have been applied to our trained model.

## 2.4 Convolutional Neural Network

Convolutional Neural Networks (CNNs) represent a biologically inspired variety of feed-forward networks, where the Connectivity between neurons tends to encapsulate the invariance of patterns to distortion or shift in given information. Convolutional Neural Networks (CNNs) typically consist of several layers, including convolutional, pooling & fully connected layers, and activation functions. The convolutional layer applies convolution operations to the input data, extracting feature probability distribution over the different classes, and enabling the network to make a decision based on this distribution as shown in Fig. 3.

### 2.4.1 Input Layer

This layer represents the initial input data, which is usually pictures in the context of applications involving computer vision. Each picture is displayed as a grid of the values of pixels (for example, RGB values). In our model the entire dataset that we resized to  $256 \times 256$ .

### 2.4.2 Convolutional Layer

This layer performs convolutional operations on the incoming data. Convolution is the process of sliding a tiny matrix known as a kernel/filter across the input data and conducting element-wise multiplication, followed by summing. This operation aids in obtaining a variety of characteristics from the supplied image. Multiple filters

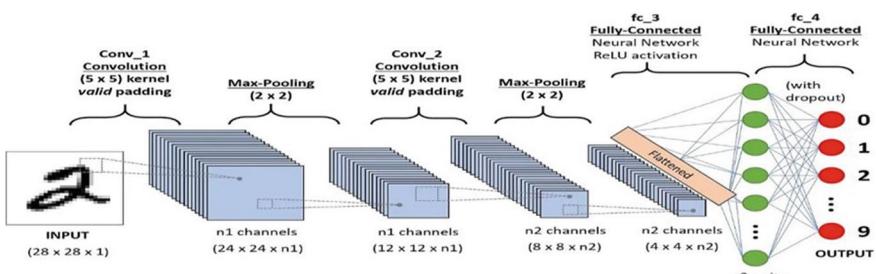


Fig. 3 Layers in convolutional neural network (CNN)

are employed in a single convolutional layer to identify various characteristics. The output of this layer is commonly known as feature maps or activation maps.

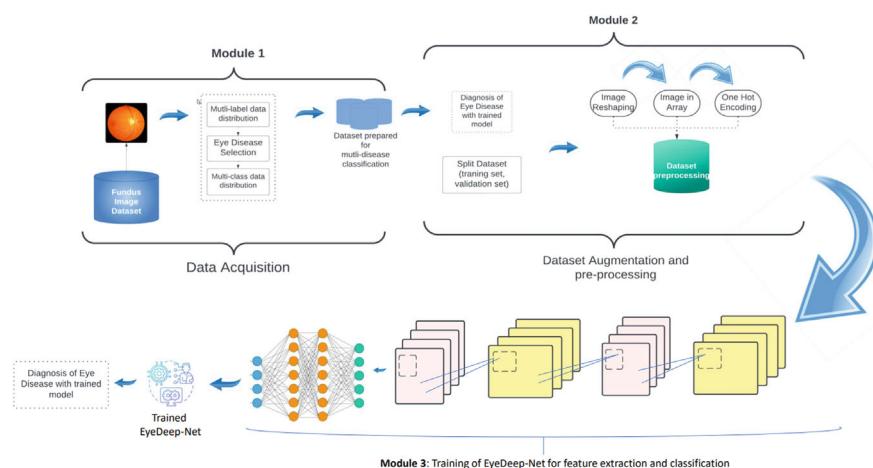
### 2.4.3 Max Pooling Layer

Pooling layers are used to minimize the spatial dimensions of feature maps while maintaining crucial information. In order for max pooling to function, the feature map is divided into rectangular, non-overlapping areas known as pooling windows or kernels. The largest value is chosen for each region and kept, with the remaining values being deleted. A new feature map with a higher degree of abstraction and a smaller spatial dimension is the end product.

## 2.5 Residual Network (ResNet18)

In this session ResNet18 introduced that a typical CNN works with a reduced no.of layers, if we increase the layer count, it might create complications associated with it quite difficult in such a way our proposed methodology as shown in Fig. 4.

Here are a few steps to see in three modules and patterns from the input through a process known as convolution. Activation functions, like the ReLU layer introduce non-linearity by replacing all negative pixel values in the feature map with zero, allowing the network to learn and model complex data effectively. The pooling layer reduces the dimensions of the convolved features, decreasing the computational cost and controlling overfitting by summarizing the features present in a region of the feature map. Fully connected layers connect each neuron in one layer to every



**Fig. 4** Workflow of the proposed architecture



**Fig. 5** Classified disease sample images

neuron in the next one, facilitating high-level reasoning. The softmax layer is applied to the o/p of a Neural Network to multiclass categorization, producing a w with that called the Vanishing/Exploding gradient. To resolve the complications of vanishing & exploding gradient, this novel architecture reacquaints the idea is Residual Blocks. We employed skip connections in this network, which skip some levels in between to connect activations of one layer to the next. Thus, after the residual block is created which is mounted to construct residual networks. The goal behind this network is to allow the network to fit the residual mapping instead of having layers learn the fundamental mapping. Adding this kind of skip connection has the benefit of allowing regularization to bypass any layer that degrades architecture performance. As a result, training a very deep neural network is achieved with the elimination of gradient issues. In this model implementation, some pre-trained images are there in the figure in which random images from the four classification images have been plotted.

## 2.6 *Proposed ResNet18 Working*

Our proposed deep CNN ResNet18 model that has been trained effectively hasn't used Pre-trained weights in this working of ResNet18. The entire dataset is split into training testing and validation datasets. ThisResNet18 trained model is optimized by using Adam optimizer trained for 100 epochs with each epoch having 28 iterations. It is monitored by training & validation loss and the model is excellently operated to achieve 100% accuracy. Here are a few sample images that we can see in Fig. 5 plotted from the dataset.

## 2.7 *Evaluation Metrics*

The execution of this ResNet18 was evaluated by using Accuracy, precision, and F1-Score. Accuracy is the ratio between the total no.of precise predictions to the overall predictions. Precision could measured by the ratio of true positive predictions to the

overall positive predictions. F1-Score is obtained by the harmonic mean of precision and sensitivity values. True Positive (TP) is a measure of the cases predicted as true to be true. True Negative (TN) is a measure of the cases predicted as false to be false. False Positive (FP) is a measure of the cases predicted as true to be false. False Negative (FN) is a measure of the cases predicted as false to be true.

### 2.7.1 Accuracy

Classification accuracy is a crucial factor in evaluating classification challenges. It represents the frequency with which the model predicts the correct result. It is calculated by dividing the number of correct predictions produced by the classifier by the total number of predictions made by the model.

$$\text{Accuracy} = \frac{TP + TN}{FN + TP + TN + FP}$$

### 2.7.2 Precision

Precision refers to the total quantity of accurate results offered by the trained model's predictions, or the percentage of accurately predicted that all the positive classes were true.

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

### 2.7.3 Recall

Recall, or sensitivity, or TPR, is the percentage of predictions that are true positive out of all true positives in the data set. It represents the model's capacity to capture all true positives. The recall metric must be maximally high.

$$\text{Recall} = \frac{TP}{FN + TP}$$

### 2.7.4 F1-Score

It is difficult to compare two models with bad recall and good recall or vice versa. Therefore, we may use F1-scores. F1-scores allow us to compare recall and accuracy simultaneously. It may be computed using the following formula.

$$F1 - Score = 2 \times Precision \times \frac{Recall}{Precision + Recall}$$

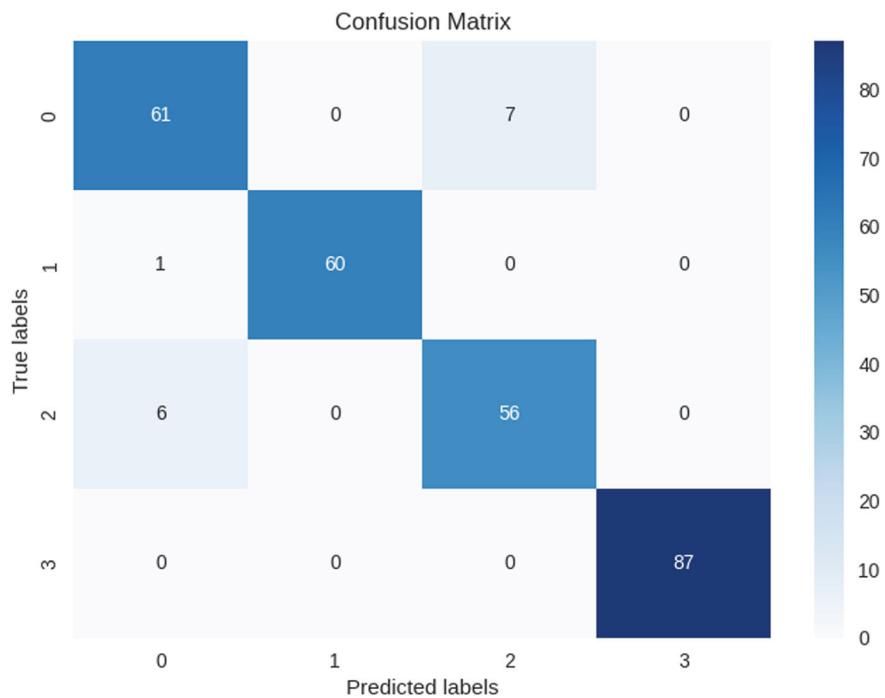
### 3 Experimental Setup

The dataset consists of Normal, Retinopathy, Cataract, and Glaucoma fundus images, where each class has approximately 1000 images. These fundus images are assembled from distinctive roots like IDRiD, HRF, Ocular recognition, etc. The dataset is downloaded from Kaggle containing 4217 images of these four labels. Given dataset is split in the ratio of 4:1 into two datasets one is the training & other is testing dataset. After splitting the dataset resize the images into 256\*256 pixels. Data augmentation is transforming the images by rotating, flipping, and cropping to increase the training dataset. Considering batch size 8 the Resnet model is trained using a training dataset. After training all batches the trained model will be tested by testing the dataset. To evaluate the performance of model metrics like accuracy, F1-score, and precision are used.

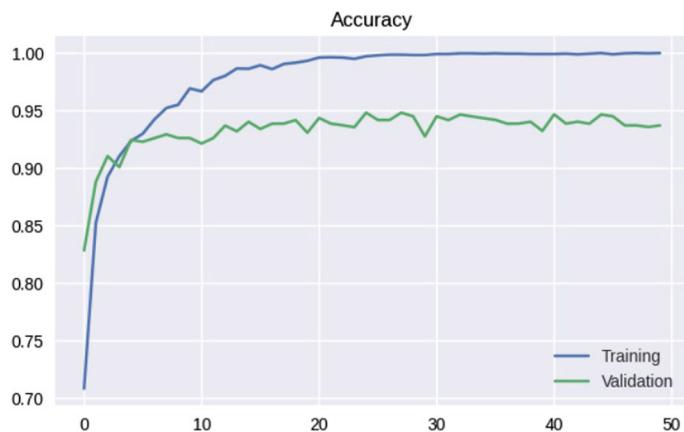
### 4 Results and Discussion

The Resnet18 model achieved remarkable performance in classifying retinal diseases, boosting a training & validation accuracy of 100 & 94%. Notably, the model excelled in discriminating the diabetic\_retinopathy and cataract classes, attaining precision and recall and F1-scores surpassing 0.99, 0.99 & 0.99 respectively. In Fig. 5 we can observe the confusion matrix of our trained model. This underscores the model's efficiency in accurately categorizing images into healthy, glaucoma, diabetic retinopathy, and cataract categories, showcasing its potential in clinical applications. In Fig. 6 our model performance can be observed by the confusion matrix. We can have clear information in the given figures and table. Figure 7 shows the performance evaluation of the Training and Validation loss function and Fig. 8 shows the Accuracy of Training and Validation dataset. In this proposed method we have considered various measurement parameters such as precision, recall, and F1-score we provided the detailed information in the Table 1.

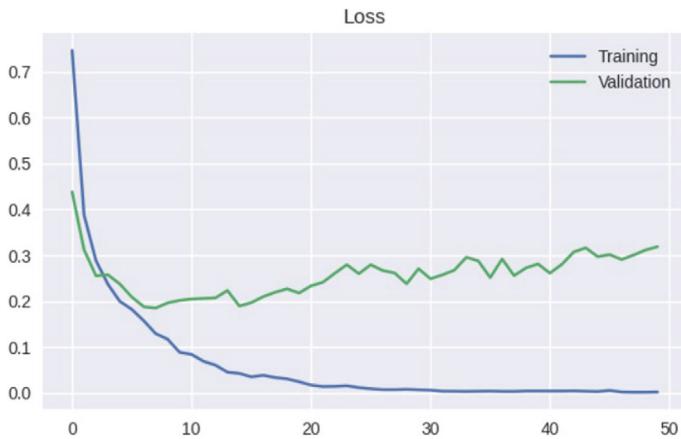
In this ResNet18 model implementation, some important modules including Data Augmentation pre-processing and data acquisition are used in the implementation of the proposed ResNet18. These modules improve the model's ability to identify retinal vessels more accurately while using fewer hyperparameters. It was first trained and evaluated on a large amount of dataset using several metrics, and it was able to segment both the arteries and veins in the input images with the minimum loss observed which is 0 loss value so that the accuracy is 100% achieved with this ResNet18. Classification results have been done effectively with this trained model.



**Fig. 6** Confusion matrix



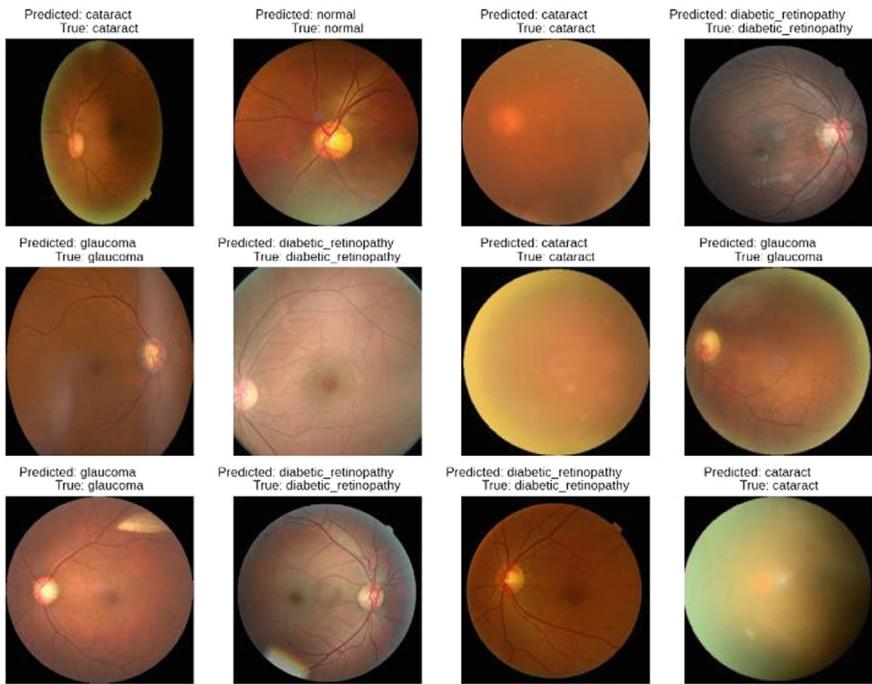
**Fig. 7** Graph between training and validation accuracy

**Fig. 8** Training and validation loss**Table 1** Performance analysis of the ResNet18 model

	Disease class	Precision	Recall	F1-score
ResNet18	Cataract	0.98	0.95	0.97
	Diabetic retinopathy	0.99	0.99	0.99
	Glaucoma	0.93	0.87	0.9
	Normal	0.86	0.94	0.9
	Accuracy	Training dataset		100%
		Testing dataset		95%
		Validation dataset		94%
ResNet50	Accuracy	72%		
ResNet101	Accuracy	65%		

Finally, we can see the classified results from our trained model in Fig. 9 shown below as we plotted for a few of them and the results consisting of all mentioned four categories. Ultimately, Table 1 presents the greatest outcomes of our ResNet18 model, outperforming other models in terms of performance.

Comparatively, we trained the ResNet50 and ResNet101 models on the same dataset, but we saw low performance from these two models ResNet101 provided an accuracy of 65%, while ResNet50 provided 72% and both models required a significant amount of training time. Thus, the results of our ResNet18 are displayed in the table.



**Fig. 9** Classification results of predicted and true labels

## 5 Conclusion

Early identification & classification of retinal disorders have paramount importance in the eye care system. Glaucoma, cataracts, and other related eye diseases have already been classified using various other earlier models. However, the proposed ResNet18 model can classify all classes of retinal diseases from the retinal images. Thus, in the field of ophthalmology, automated vessel detection, and categorization is essential. This study presents a novel ResNet18, delivering excellent results with 100% accuracy, 99% precision, Recall, and F1-Score respectively with minimum Loss value. In the final note, ResNet18 is an efficient model for the classification of various eye diseases.

## 6 Future Scope

In some circumstances, ResNet18 may be at risk for overfitting, especially when working with tiny datasets or when model capacity exceeds what is required for the job at hand. To minimize this problem, regularization techniques or model architecture tweaks may be necessary. Limited interpretability is introduced. While ResNet-18's skip connections can help with interpretability to some extent, deep neural networks in general are sometimes criticized for their lack of interpretability, since it can be difficult to grasp how the network makes decisions, particularly at deeper layers. The detection and evaluation of retinal diseases have exciting potential in store. These include the development of non-invasive monitoring technologies for real-time tracking of disease progression, the identification of quantitative biomarkers for objective disease severity assessment, the potential for remote screening to reach underserved populations, the emergence of personalized treatment approaches based on genetic profiling and individualized medicine, and the integration of advanced machine learning and artificial intelligence for early disease detection. Future developments hold promise for improving disease monitoring, early diagnosis, and individualized treatment plans, with the ultimate goal of protecting and promoting visual health for those who are susceptible to retinal disorders.

**Acknowledgements** The authors of this research article thank the Science and Engineering Research Board (SERB), India for granting this work under the Core Research Grant (CRG) funding scheme of Science Engineering and Research Board (SERB) under grant no—CRG/2023/005474. The authors also express their gratitude and thanks to the team of Center for Medical Imaging Studies (CMIS) for their support in this work.

**Conflict of Interest** There is no conflict of interest between the authors.

## References

1. Weni I, Utomo PEP, Hutabarat BF, Alfallah M (2021) Detection of cataract based on image features using convolutional neural networks. *Indones J Comput Cybern Syst* 15(1):75–91
2. Mishra S, Hanchate S, Saquib Z (2020) Diabetic retinopathy detection using deep learning. In: Proceedings of the 2020 international conference on smart technologies in computing, electrical and electronics (ICSTCEE). IEEE, pp 1–13. <https://doi.org/10.1109/ICSTCEE49637.2020.9277506>
3. Ghazal M, Mahmoud AH, El-Baz A (2020) Accurate detection of non-proliferative diabetic retinopathy in optical coherence tomography images using convolutional neural networks. <https://doi.org/10.1109/ACCESS.2020.2974158>
4. Toki SA, Rahman S, Fahim SMB, Mostakim AA (2022) RetinalNet-500: a newly developed CNN model for eye disease detection. In: Proceedings of the 2nd international mobile intelligent and ubiquitous computing conference (MIUCC)
5. Hassan SAE, Akbar S et al (2021) Deep learning-based automatic detection of central serous retinopathy using optical coherence tomographic images. In: Proceedings of IEEE conference on computer-aided industrial design & applications (CAIDA). IEEE, pp 1–10. <https://doi.org/10.1109/CAIDA51941.2021.9425161>

6. Veena HN, Muruganandham A, Kumaran TS (2022) A novel optic disc and optic cup segmentation technique to diagnose glaucoma using deep learning convolutional neural network over retinal fundus images. *J King Saud Univ-Comput Inf Sci* 34(8):6187–6198
7. Yaqoob MK, Ali SF, Bilal M, Hanif MS, Al-Saggaf UM (2021) ResNet based deep features and random forest classifier for diabetic retinopathy detection. *Sensors* 21(11):3883. <https://doi.org/10.3390/s21113883>
8. Sattigeri SK, Harshith N, Gowda DN, Ullas KA, Aditya MS (2022) Eye disease identification using deep learning. *Int Res J Eng Technol (IRJET)* V9(7):101–109. <https://www.irjet.net/archives/V9/I7/IRJET-V9I7185.pdf>
9. Bhatkar AP, Kharat GU (2015) Detection of diabetic retinopathy in retinal images using MLP classifier. In: Proceedings of the 2015 IEEE international symposium on nanoelectronic and information systems (IESNIS). IEEE, pp 1–2
10. Pratap T, Kokil P (2019) Computer-aided diagnosis of cataract using deep transfer learning. *Biomed Signal Process Control* 53:101916
11. Vasudeva J, Rastogi P, Yadav J (2020) Glaucoma detection and segmentation as computer-aided design: a review and study. In: Proceedings of the 2020 IEEE 17th India council international conference (INDICON). IEEE, pp 1–10
12. Mustafa H, Ali SF, Bilal M, Hanif MS (2022) Multi-stream deep neural network for diabetic retinopathy severity classification under a boosting framework. *IEEE Access* 10:113172–113183
13. Jain L, Murthy HVS, Patel C, Bansal D (2020) Retinal eye disease detection using deep learning. In: Proceedings of the fourteenth international conference on information processing (ICINPRO). <https://doi.org/10.1109/ICINPRO43533.2018.9096838>

# LSTM-Based Portfolio Optimization with Gerber Covariance Estimator for Increased Robustness



Ishita Mehta and Kartik Gupta

**Abstract** Traditional methods of optimizing investment portfolios often rely on rigid models that are not able to fully capture the complexity and unpredictability of financial markets. Previous research has primarily centered around statistical analysis, which struggles to capture the complex interdependencies between data points that ML-based models may be able to. In this paper, we have constructed a hybrid model incorporating machine learning and statistical methods to propose a dynamic and data-driven approach to optimize the portfolio using a Long Short-Term Memory (LSTM) network and a covariance matrix based on the Gerber statistic. It estimates the return prices and consequently constructs a Gerber estimator covariance matrix. This yields two-fold advantages, added robustness and improved risk management. Our study utilizes S&P500 tech stocks to evaluate the model and constructs 3 different portfolios for asset allocation, i.e., Minimum CVaR, Mean Risk, and Hierarchical Risk Parity. The annualized Sharpe Ratio for HRP achieved is 1.75, and the metrics show effective risk management. The model is unique due to its hybrid nature which combines the data-centricity and the flexibility of LSTM to capture any non-linear patterns which is capable of adjusting to evolving market conditions. By leveraging the statistical rigor of the covariance matrix to yield promising results, the model can gauge the stochastic nature of financial markets to a reasonable extent, thereby helping to maximize returns and minimize risks.

**Keywords** Portfolio optimization · LSTM · Gerber statistic · Hierarchical clustering · Covariance matrix

---

I. Mehta (✉) · K. Gupta

Department of Software Engineering, Delhi Technological University, New Delhi, Delhi 110042, India

e-mail: [ishitamehta028@gmail.com](mailto:ishitamehta028@gmail.com)

K. Gupta

e-mail: [kartikgupta\\_se20b8\\_51@dtu.ac.in](mailto:kartikgupta_se20b8_51@dtu.ac.in)

## 1 Introduction

While the “Efficient Market Hypothesis” remains a cornerstone theory in finance and economics, advocating the inherent difficulty in forecasting financial asset prices, empirical evidence from numerous studies suggests a nuanced reality. Many studies have proved that stock prices and returns are, to some extent, predictable. One of the most essential concerns in asset management is portfolio optimization [1]. Optimize allocation to maximize returns across a group of assets with varying yields given a fixed quantity of available resources. The end goal of optimizing portfolios is to find the best balance between risk and return over a certain time period [2].

In the famous Modern Portfolio Theory (MPT), Markowitz explicitly articulated the problem. The Markowitz model, which chooses the allocation of investors’ investments based on a mean–variance analysis [3], is the foundation of portfolio optimization theory. Many statistical techniques are also used. Generalized autoregressive conditional heteroscedasticity (GARCH) [4], Markowitz mean–variance, and Autoregressive integrated moving average (ARIMA) [5], among others are examples. A substantial amount of work has since been added to the study of portfolios by employing a range of different methodologies, including neural networks [6], evolutionary algorithms [7], Random Matrix Theory (RMT) filtering [8, 9], and K-means hierarchical clustering [10]. Machine learning is also popularly used [11, 12], a notable example is the XGBoost [13] which improves performance while taking into account risk.

Due to the highly stochastic nature of financial markets, it is difficult to predict the investment return of financial assets with a certain level of precision. Deep learning models these days are performing better than other statistical methods and time-series models [14]. However, there is a gap between how well models generalize out-of-sample data in real-world and how robust these models are when deviated from certain assumptions from the market dynamics. Ensuring risk diversification in the allocation of portfolio optimization using dynamic methods is imperative to ensure robust asset allocation [15]. Errors in parameter estimations result in inaccurate asset allocation and are therefore not optimized. Parameter optimization and hyperparameter tuning are significant research gaps for feature selection and fine-tuning the model and add to robustness of the model.

In this paper, we propose a hybrid architecture that predicts the closing prices from an LSTM model and uses that as an input for the Geber Covariance matrix for portfolio construction [16, 17]. The weights generated by the matrix help optimize the portfolio. The weight values decide the asset allocation strategy. The portfolio construction in this paper is a panel of 3 varied portfolios for the purpose of drawing comparisons. These are—Minimum CVaR [18], Maximum Sharpe Ratio [19, 20] and Hierarchical Risk Parity (HRP) [21]. The mean-risk portfolio seeks to optimize the tradeoff between the expected return and the risk in the portfolio. The HRP portfolio employs hierarchical clustering techniques to group assets based on their correlation structure and allocates weights to clusters and individual assets within clusters to achieve risk parity.

The contributions by the authors are as follows—Conceived, designed, technical analysis, and written by Ishita Mehta. Written, formatted, and technical assistance by Kartik Gupta.

The rest of the paper is structured as follows. We lay out the methodology and implementation of our proposed model. We then discuss the results and compare the panel of portfolios against baseline algorithms. Finally, we summarize our findings and conclude with future scope.

## 2 Methodology

In this section, we discuss the functionality of each step in the framework. We first setup the Gerber Covariance matrix and the Sharpe Ratio. For the purpose of our analysis, we have used a panel of three different portfolios, each has a different optimization goal with different constraints.

### 2.1 Gerber Covariance Estimator

The Gerber Covariance Matrix [16, 17] uses robust statistical methods and outlier detection to improve accuracy, especially with noisy data and irregular data. Extending idea from Kendall’s Tau, the Gerber statistic considers the proportion of simultaneous co-movements in series, particularly when their amplitudes exceed data-dependent thresholds. Unlike traditional covariance matrices, which are susceptible to the influence of outliers and non-Gaussian distributions, like the Ledoit-Wolf shrinkage method [22], the Gerber Covariance Matrix employs robust estimation techniques to mitigate the impact of extreme observations by ignoring the fluctuations below a set threshold and limiting the effects of extreme movements.

Let the number of assets in our index be  $n$ , the return on each asset  $r_{i,t}$ , time-period  $t \in [1, T]$ ,  $\sigma$  be sample standard deviation of historical asset returns, the threshold value,  $c \in [0, 1]$ . Then, the Gerber Statistic between asset  $i$  and asset  $j$ ,  $g_{ij}$  is defined as

$$g_{ij} = \frac{n_{ij}^c - n_{ij}^d}{n_{ij}^c + n_{ij}^d} \quad (1)$$

where  $n_{ij}^c$  are the number of concordant pairs for time-series  $i, j$ , and  $n_{ij}^d$  are the number of discordant pairs.

Consequently, the Gerber Covariance Matrix is defined by

$$\left( \sum_G \right)_{ij} = g_{ij} \sigma_i \sigma_j, \quad i \in [1, n], j \in [1, n] \quad (2)$$

## 2.2 Sharpe Ratio

The risk-adjusted return of an investment portfolio can be measured using the Sharpe Ratio. The excess return per unit of risk is quantified while considering the volatility of the investment. The higher the value of Sharpe Ratio better is the performance with risk taken into account. The formula for the Sharpe Ratio (S) is

$$S = \frac{R_p - R_f}{\sigma_p} \quad (3)$$

where  $R_p$  is the expected return from the portfolio,  $R_f$  is the rate of return which is considered risk-free, and  $\sigma_p$  is the standard deviation (volatility) of the portfolio's return.

## 2.3 Objective Function

The risk-adjusted return of an investment portfolio can be measured using the Sharpe Ratio. The excess return per unit of risk can be quantified, considering the volatility of the investment.

**Minimum CVaR.** This is a convex portfolio prioritizes risk reduction instead of optimizing on return values. In our study, the objective function is set to Minimize Risk and the risk parameter is set to Conditional Value-at-Risk (CVaR).

$$\begin{aligned} & \text{Minimize} && risk_i(w) \\ & \text{S.T} && w^T \mu \geq \min\_return \\ & && Aw \geq b \\ & && risk_j(w) \leq max\_risk_j \quad \forall j \neq i \end{aligned} \quad (4)$$

where  $w$  is the vector of assets weights,  $\mu$  is the vector of assets' expected returns,  $A$  is the matrix representing constraints on portfolio weights, matrix  $b$  specifies constraint boundaries.

**Maximum Sharpe Ratio.** This is a convex portfolio and uses the Mean-Risk optimization strategy to maximize the Sharpe Ratio by optimizing on expected return. In our study, the objective function is set to Maximize Ratio with Risk parameter set to Variance.

$$\begin{aligned} & \text{Minimize} && \frac{(w^T \mu - r_f)}{risk(w)} \\ & \text{S.T} && risk(w) \leq max\_risk_i \\ & && w^T \mu \geq min\_return \\ & && Aw \geq b \\ & && risk_j(w) \leq max\_risk_j \quad \forall j \neq i \end{aligned} \quad (5)$$

**Hierarchical Risk Parity.** This method utilizes hierarchical clustering and seriation to organize assets[23], followed by recursive bisection to divide clusters into sub-clusters. Total cluster risk is computed for each sub-cluster using inverse-risk allocation, and a weighting factor is derived to update cluster weights.

$$\begin{aligned} & \text{Minimize} && \text{risk}_i(w) \\ \text{S.T.} & \quad \tilde{\alpha}_i = \max \left[ \min \left[ \alpha_i, \sum_i \alpha_i^{-\max} \right], \sum_i \alpha_i^{-\max} \right] \end{aligned} \quad (6)$$

where  $\alpha_i^{-\max}$ ,  $\alpha_i^{-\min}$  are respective  $1 \times N$  constraint vectors for clustering.

$$\text{and split factor } \alpha_i = \frac{\left[ V_i^{\sim j} \right]^{-1}}{\sum_j \left[ V_i^{\sim j} \right]^{-1}} \quad (7)$$

where  $V_i^j$  is the covariance matrix of elements within cluster  $j$ .

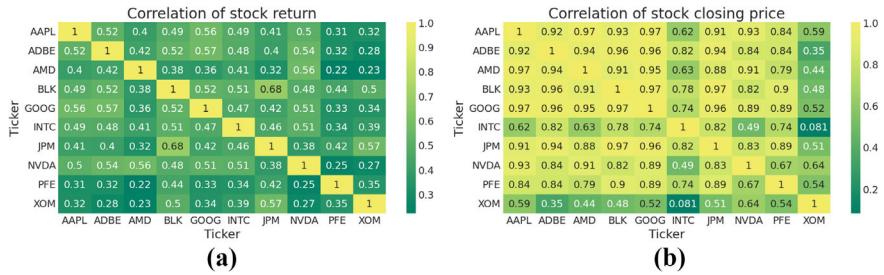
### 3 Implementation

The models and optimization functions described in the previous section serve as the basis for our experimental setup. In this section, we layout the working mechanism of our model and layout the pipeline and use comparative analysis for assessment.

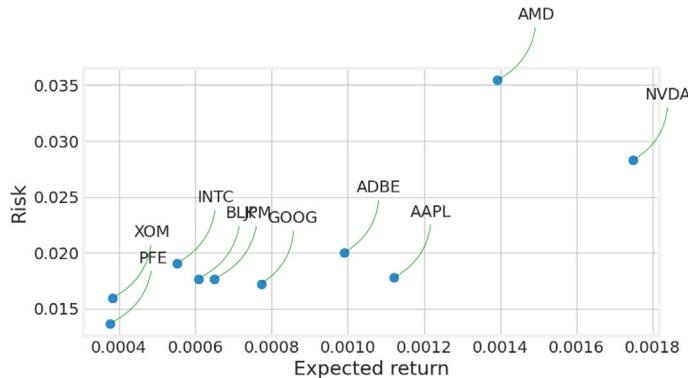
#### 3.1 Dataset

We began by collecting historical stock data from Yahoo! Finance, encompassing the S&P 500 index and eight prominent technology stocks from January 1, 2010 to January 1, 2024. These are stocks from Apple, Google, Nvidia, Adobe, Intel, AMD, JPMorgan Chase & Co, Blackrock, Exxon Mobil, Pfizer, denoted by [AAPL, GOOG, NVDA, ADBE, INTC, AMD, JPM, BLK, XOM, PFE]. The dataset contains daily closing values spanning a time period of 14 years. The dataset underwent preprocessing to handle missing values, and any inconsistencies to ensure data integrity and consistency, then accounted for stock splits and adjusted closing prices accordingly. The validation dataset accounts for 1/3rd of the total dataset.

The time-series data was analyzed by visualizing the correlation matrix of stocks returns and closing prices. As observed in Fig. 1, some stocks like Blackrock and JP Morgan are strongly correlated with each other while some have very less correlation like Nvidia and Exxon Mobil. This is due to the difference in their respective industries.



**Fig. 1** Correlation matrix of **a** Stock returns and **b** Stock closing prices

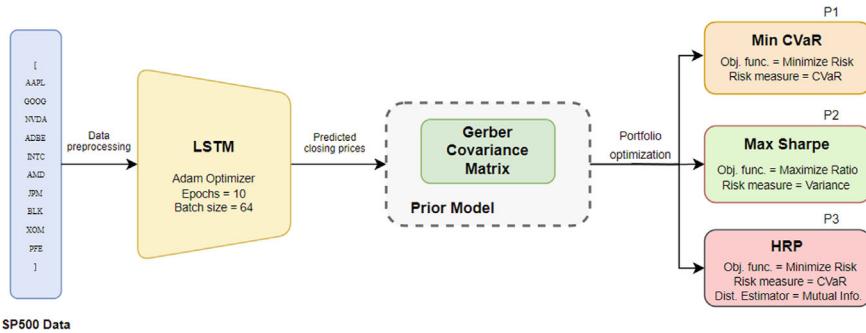


**Fig. 2** Risk versus expected returns for tickers

The risk versus expected returns for each ticker is visualized in Fig. 2. A proportion of stocks have a moderate risk-return ratio. Notable exceptions are AMD and Nvidia, offering a higher risk-return ratio, with potential of higher rewards, making them attractive investment options for investors.

### 3.2 Model Architecture

The dataset with our chosen stocks from SP500 is preprocessed and normalized. Input–Output pairs are generated spanning 60-time steps to predict subsequent points and passed into the LSTM. The LSTM architecture comprises two layers—a 128-unit layer followed by a 64-unit layer for temporal dependency capture, followed by dense layers for mapping. We have used the Adam Optimizer and iterated over 10 epochs with a batch size of 64. The validation data is used to check the accuracy of the predicted values and a Root Mean Squared (RMSE) is calculated for each stock. The predicted stock closing prices generated from our LSTM model form the basis



**Fig. 3** Visual representation of the hybrid model. As shown, the predicted LSTM data is passed for statistical portfolio construction

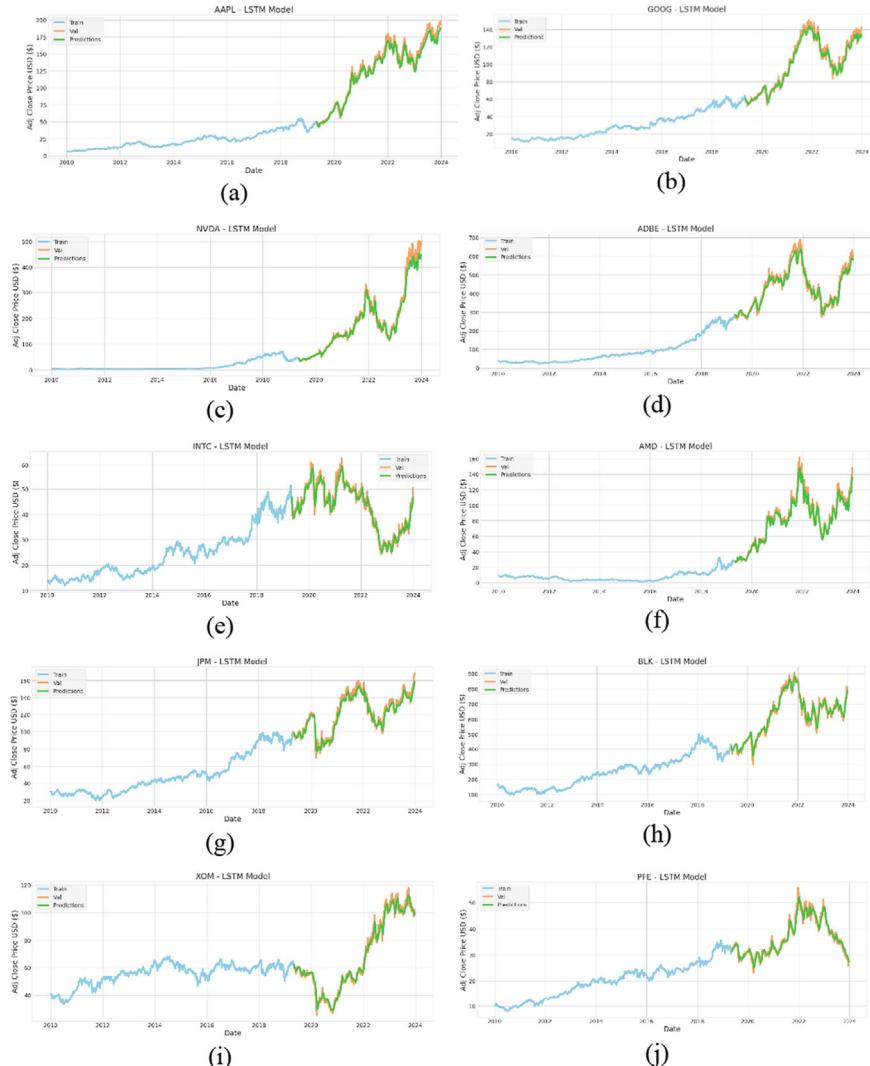
for the input for our statistical analysis for portfolio construction and optimizing portfolio weights.

Before portfolio construction and optimization, a Prior Model is constructed. A Prior Model represents our assumptions about the model that is used to estimate the distribution of asset returns. Our assumptions about the expected returns and their distribution, a covariance matrix and the assets return estimation distribution make up the Prior Model. For our analysis, we have chosen the Gerber covariance matrix with the James–Stein shrinkage for the estimation of expected returns. This model is used to fit the selected portfolio strategy and optimized based on their respective objective functions, subjected to certain constraints.

Our chosen panel of portfolio optimizers are Max Sharpe Ratio, Min CVaR, and Hierarchical Risk Parity (HRP). These employ different strategies to optimize portfolio weights and generate different stock distributions of portfolio compositions. These methodologies have been applied to generate predicted cumulative returns over our testing period spanning from July 2019 to July 2023. The model overview is illustrated in Fig. 3.

## 4 Results

In this section, we present the results of our experimental setup. Figure 4 depicts the predicted closing prices and is contrasted with the validation closing prices. Overall, the LSTM is quite effective in predicting the closing prices as demonstrated by the close alignment of validation values of the test set and our predicted values. RMSE values for our chosen dataset of stocks are presented in Table 1.



**Fig. 4** LSTM-predicted and validation closing prices of **a** AAPL **b** GOOG **c** NVDA **d** ADBE **e** INTC **f** AMD **g** JPM **h** BLK **i** XOM **j** PFE

#### 4.1 Portfolio Optimization Using Gerber Estimator

LSTM prices fed into the Gerber covariance for prior model generation are used for portfolio construction and optimization. The cumulative returns of our portfolios, taken from July 2019 to July 2023 are visualized in Fig. 5. All portfolios perform well with only the period around March 2020 showcasing a severe dip in returns which is owed to the effect of COVID-19 in markets. Evidently, Max Sharpe offers

**Table 1** RMSE values for LSTM-predicted closing prices

Stock	Company	RMSE value
AAPL	Apple	6.37251
GOOG	Google	4.49360
NVDA	Nvidia	18.77117
ADBE	Adobe	23.08293
INTC	Intel	1.99042
AMD	AMD	5.65733
JPM	JPMorgan chase & Co	5.13491
BLK	Blackrock	24.64846
XOM	Exxon mobil	2.93670
PFE	Pfizer	1.45688



**Fig. 5** Cumulative returns of Min. CVaR, mean risk, and HRP portfolios using gerber covariance estimator

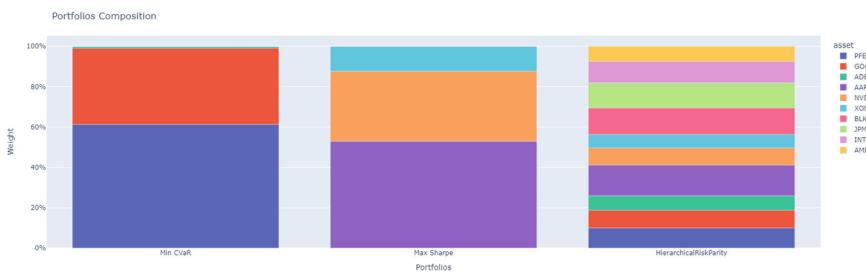
the highest cumulative returns, followed by HRP portfolio and finally Min CVaR portfolio.

Table 2 presents the weight distribution across these portfolios and the same can be visualized in Fig. 6.

Metric analysis was conducted to evaluate portfolio stability and performance across varying market conditions and are presented in Table 3. Evidently, the Max Sharpe portfolio shows superior risk-adjusted returns and the risk-return tradeoff is well optimized. Conversely, the Min CVaR portfolio exhibits a conservative risk management approach with lower risk measures such as maximum drawdown and VaR values. The HRP portfolio is well-diversified, indicated by its effective number of assets and has the highest annualized Sharpe ratio out of the panel and has a balanced risk allocation.

**Table 2** Weight distribution of stocks in different portfolios

Index	Asset	Weight (%)
Min CVaR	ADBE	1
	GOOG	38
	PFE	61
Max sharpe	AMD	3
	JPM	4
	GOOG	9
	XOM	16
	NVDA	26
	AAPL	42
HRP	AMD	7
	XOM	7
	ADBE	7
	NVDA	9
	GOOG	9
	PFE	10
	INTC	11
	JPM	12
	BLK	13
	AAPL	15



**Fig. 6** Portfolio composition of stocks using gerber covariance matrix **a** Min CVaR **b** Max sharpe **c** Hierarchical risk parity

**Table 3** Evaluation metrics for the panel of portfolios

Index	Sharpe ratio	Sortino ratio	Annualized sharpe ratio	Effective number of assets	Calmar ratio	Max drawdown	VaR at 95%	CVaR at 95%
Min CVaR	0.26	0.37	0.42	1.928	0.00093	42.04%	2.32%	3.37%
Max sharpe	0.79	0.11	1.75	2.402	0.0045	37.39%	3.26%	4.79%
HRP	0.58	0.85	1.28	9.312	0.0024	41.25%	2.46%	3.92%

## 4.2 Comparison with Baseline Algorithms

The cumulative results and metrics generated from our panel of portfolios significantly surpass previous methodologies. Our hybrid model exhibits superior performance compared to traditional statistical methods like Mean–Variance, ARIMA, and the Capital Asset Pricing Model (CAPM), and gives competitive results with other ML-based models such as XGBoost and Random Matrix Theory Filtering, and evolutionary algorithms. Utilizing the Gerber covariance estimator shows promising results in handling extreme market fluctuations and adds robustness to portfolio performances.

## 5 Conclusion

In this paper, we presented a model that optimizes portfolio with different optimization functions using predicted prices from an LSTM that were fed into the Gerber covariance matrix for prior model estimation. This pipeline has demonstrated a significant improvement in performance and error-rate compared to traditional methods such as Mean–Variance model. Utilizing an LSTM also helps in adapting to changing market conditions when compared to their traditional counterparts.

We compare our method to a wide range of popular algorithms that form the basis for portfolio management and optimization. Our testing period ranges from 2010 to 2024. The paper's findings suggest that these models can enhance the adaptability of portfolios, instead of the traditional statistical models employed which can be better at managing risk and returns for investors and corporations. The results show that our model performed well on set benchmarks and suggests that ML-based learning models are highly adaptable and resilient at handling real-world markets and are a powerful tool that investors can use in practical markets and real-world scenarios.

In subsequent works in the future, we aim to advance our model by integrating sentiment analysis into the model architecture. Psychological and behavioral aspects are rarely taken into consideration while modeling market scenarios. Market and investor sentiments significantly impact the market dynamics and therefore the prices of financial instruments. This leaves a wide research gap and scope to use sentiment analysis in optimal asset allocation. Additionally, we aim to incorporate alternate risk measures such as tail or downside risk for a more nuanced and effective risk management and advance our model with improved generalization capabilities to enhance out-of-sample performance. To adapt to evolving market conditions, real-time portfolio adjustment methods that utilize dynamic ML algorithms or ensemble methods will further improve the overall performance metrics.

## References

1. Jorion P (1992) Portfolio optimization in practice. *Financ Anal J* 48(1):68–74. <https://doi.org/10.2469/faj.v48.n1.68>
2. Durall R (2022) Asset allocation: from Markowitz to deep reinforcement learning. *SSRN Electron J.* <https://doi.org/10.2139/ssrn.4148379>
3. Sharpe WF, Markowitz HM (1989) Mean-variance analysis in portfolio choice and capital markets. *J Finance* 44(2):531. <https://doi.org/10.2307/2328607>
4. Sahamkhadam M, Stephan A, Östermark R (2018) Portfolio optimization based on GARCH-EVT-copula forecasting models. *Int J Forecast* 34(3):497–506. <https://doi.org/10.1016/j.ijforecast.2018.02.004>
5. Mendes RRA, Paiva AP, Peruchi RS, Balestrassi PP, Leme RC, Silva MB (2016) Multiobjective portfolio optimization of ARMA–GARCH time series based on experimental designs. *Comput Oper Res* 66:434–444. <https://doi.org/10.1016/j.cor.2015.05.001>
6. Ma Y, Han R, Wang W (2021) Portfolio optimization with return prediction using deep learning and machine learning. *Expert Syst Appl* 165. <https://doi.org/10.1016/j.eswa.2020.113973>
7. Cura T (2009) Particle swarm optimization approach to portfolio optimization. *Nonlinear Anal Real World Appl* 10(4):2396–2406. <https://doi.org/10.1016/j.nonrwa.2008.04.023>
8. Sharifi S, Crane M, Shamaie A, Ruskin H (2004) Random matrix theory for portfolio optimization: a stability approach. *Physica A* 335(3–4):629–643. <https://doi.org/10.1016/j.physa.2003.12.016>
9. Bai Z, Liu H, Wong W (2009) Enhancement of the applicability of Markowitz's portfolio optimization by utilizing random matrix theory. *Math Financ* 19(4):639–667. <https://doi.org/10.1111/j.1467-9965.2009.00383.x>
10. Raffinot T (2017) Hierarchical clustering-based asset allocation. *J Portf Manag* 44(2):89–99. <https://doi.org/10.3905/jpm.2018.44.2.089>
11. Ban G-Y, El Karoui N, Lim AEB (2018) Machine learning and portfolio optimization. *Manage Sci* 64(3):1136–1154. <https://doi.org/10.1287/mnsc.2016.2644>
12. 8 Machine Learning Optimization Algorithms & Portfolio Allocation (2020)
13. Dezhkam A, Manzuri MT (2023) Forecasting stock market for an efficient portfolio by combining XGBoost and Hilbert–Huang transform. *Eng Appl Artif Intell* 118:105626. <https://doi.org/10.1016/j.engappai.2022.105626>
14. Zhang Z, Zohren S, Roberts S (2020) Deep learning for portfolio optimization. *J Financ Data Sci* 2(4):8–20. <https://doi.org/10.3905/jfds.2020.1.042>
15. Zhang Y, Su Y, Liu W, Yang X Portfolio optimization with LSTM-based return and risk information. [Online]. <https://ssrn.com/abstract=4215299>
16. Gerber S, Markowitz HM, Ernst PA, Miao Y, Javid B, Sargent P (2021) The Gerber statistic: a robust co-movement measure for portfolio optimization. [Online]. [www.stat.rice.edu/](http://www.stat.rice.edu/)
17. Smyth W, Broby D (2022) An enhanced gerber statistic for portfolio optimization. *Financ Res Lett* 49. <https://doi.org/10.1016/j.frl.2022.103229>
18. Alexander S, Coleman TF, Li Y (2006) Minimizing CVaR and VaR for a portfolio of derivatives. *J Bank Finance* 30(2):583–605. <https://doi.org/10.1016/j.jbankfin.2005.04.012>
19. Vinzelberg A, Auer BR (2022) A comparison of minimum variance and maximum sharpe ratio portfolios for mainstream investors. *J Risk Financ* 23(1):55–84. <https://doi.org/10.1108/JRF-02-2021-0021>
20. Bailey DH, Lopez de Prado MM (2011) The sharpe ratio efficient frontier. *SSRN Electron J.* <https://doi.org/10.2139/ssrn.1821643>
21. López De Prado M Building diversified portfolios that outperform out-of-sample
22. Ledoit O, Wolf M (2017) Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *Rev Financ Stud* 30(12):4349–4388. <https://doi.org/10.1093/rfs/hhx052>
23. Lohre H, Rother C, Schäfer KA (2020) Hierarchical risk parity: accounting for tail dependencies in multi-asset multi-factor allocations. *Mach Learn Asset Manag.* Wiley, pp 329–368. <https://doi.org/10.1002/9781119751182.ch9>

# Handwritten Character Recognition from Small Grayscale Images Using Pre-trained Models



D. Manibharathi , C. Vasanthanayaki, and Sanjeev Kumar

**Abstract** Handwritten Character Recognition (HWCR) requires more features to work with handwritten characters because of their variable and non-uniform nature. Challenges in HWCR increased further when the number of classes increased and mathematical symbols were included. The main aim of this research work is to adapt the pre-trained models to work with multiclass classification for single channel grayscale images of mathematical symbols along with handwritten alphanumeric characters and compare their performances. A customized CNN layer was introduced to convert the single-channel grayscale images to three channels as a preliminary step in recognizing handwritten characters. After introducing the customized CNN layer, various pre-trained models were used as hidden layers to compare the performance of different pre-trained models in recognizing characters in handwritten grayscale images of small size ( $32 \times 32$ ). The separate datasets containing English capital letters, digits and mathematical symbols were merged as a single dataset of 42 classes for this study. EfficientNetB0, ResNet50, MobileNetV2, VGG16, DenseNet121 and conventional CNN models were employed to recognize the character. VGG16, DenseNet121, and CNN were very fast in recognizing the character in the grayscale images. These three models reached about 90% accuracy at the first epoch itself. The test accuracy of these three models was around 96% and higher than the others. Conventional CNN took less time than other models. Conventional VGG16 was better than other models in terms of accuracy.

**Keywords** Handwritten characters recognition · Grayscale images · Pre-trained models · Mathematical symbols

---

D. Manibharathi   
Government College of Engineering, Salem, India  
e-mail: [manibharathi@gcesalem.edu.in](mailto:manibharathi@gcesalem.edu.in)

C. Vasanthanayaki  
Government College of Engineering, Bodinayakkanur, India

S. Kumar  
University of Illinois, Urbana-Champaign, USA  
e-mail: [sanjeev5@illinois.edu](mailto:sanjeev5@illinois.edu)

## 1 Introduction

Optical Character recognition (OCR) has emerged as an important research field since the early days of computer science and engineering development. It is widely used for extracting and interpreting the characters included in scanned or printed data.

Due to the limited set of symbols that are available as fonts in computer systems and the uniformity in character, word and sentence placement, and size, recognition of machine printed documents has evolved quicker and sooner.

However, handwritten character recognition (HWCR) remains challenging because of the need for more features to work with handwritten characters, texts because of their variable and non-uniform nature. Greater variability of writing styles, resulting in virtually infinite ways of writing the same symbol. The similarity between symbols are high. Touching and overlapping characters also increases the difficulty.

Other than these the poor image quality, noise, degradation of documents have more effect on handwritten documents when compared with printed documents [1].

The methods of obtaining handwritten character information are offline recognition and online recognition. In ‘offline recognition’ the pre-existing handwritten documents are acquired via scanners or photo cameras [2]. In this the images store the characters information as the image intensity, that is, the values of each pixel in the coded image [3]. In ‘online recognition’ the writing is directly made in devices capable of capturing the coordinates of the writing location and other information, such as stroke velocity, pen pressure or the order of the traces. Usage of temporal data for online recognition yields better results.

Because of the high performance of deep neural networks (DNNs), particularly convolutional neural networks (CNNs) [4–6], for several image recognition tasks, the DNN architecture is expected to demonstrate promising performance for problems involving handwritten characters. Some attempts to introduce DNNs for alphabets, digits, and mathematical symbol recognition have been proposed [7, 8]. Pretrained models offer advantages in deep learning applications [9, 10], like in image and character recognition, since they are efficient in terms of time and resources, require less data, and can prevent overfitting. By utilizing a larger dataset, they have acquired knowledge that has led to the development of more generalized characteristics. They enhance performance by utilizing patterns acquired from a more extensive dataset and serve as an initial reference for transfer learning, enhancing the training process’s efficiency.

Additionally, they function as standards for novel models to assess their effectiveness in comparison to well-established models. Bhati and Garg compared the performance of two CNN architectures, VGG16 and DenseNet121, with transfer learning for the handwritten Devanagari character dataset [11]. Their findings indicated that the DenseNet121 model with deep fine-tuning method outperformed the VGG16 model. Chandure and Inamdar (2023) created an image dataset for handwritten MODI characters and used a supervised Transfer Learning approach utilizing a pre-trained Deep Convolutional Neural Network, specifically Alexnet, for feature

extraction. The extracted features were then used to train a Support Vector Machine for classification [12]. Albatnah and colleagues introduced a method for intelligent recognition of Arabic handwriting using different independent and hybrid convolutional neural network (CNN) architectures [13]. Several hybrid models, including deep learning, have been constructed. The approach employed in this case involved integrating deep learning and machine learning techniques to create hybrid models capable of extracting features and performing classification. Among the two trial datasets, the deep learning model applied to the MNIST dataset yielded the most favorable outcomes, particularly the transfer-learning model.

Grayscale image datasets such as MNIST, EMNIST, IAM, and synthetic datasets are widely used in OCR and can be accessed through academic sources and repositories like Kaggle. When these datasets are used, necessary preprocessing of the images is required, such as resizing them to a consistent dimension and channels, normalizing pixel values, and converting them to the appropriate format for training. Numerous difficulties emerge as the total number of output classes in a classification task increase. These factors encompass increased model intricacy, imbalanced data, training difficulty, increased risk of overfitting, diminished interpretability, lowered performance, and increased labeling effort. The intricacy of the model may necessitate larger neural network structures, which might be computationally expensive and demand more data. Imbalanced data might result in skewed models, posing challenges in addressing the issue. Also, it can be time-consuming, and the model may end up memorizing training instances instead of being able to adapt appropriately to new, unseen data. From the above literature, the following research gaps were identified. (1) Challenge in recognizing characters when number of classes increases (2) Mathematical symbols were not explored much (3) Pre-trained models are trained with color images.

The scope of the research includes the exploration of HWCR, particularly from small grayscale images of alphanumeric and mathematical symbols. It involves analyzing the effectiveness of various pre-trained models in this context.

The objective of the research work was to recognize handwritten characters from small grayscale images of alphanumeric and mathematical symbols using pre-trained models and compare their performances. The rest of the sections of the paper were organized as follows. The second section describes the methods adapted in this research work. The third section describes the results and discussion. The final section describes the conclusion.

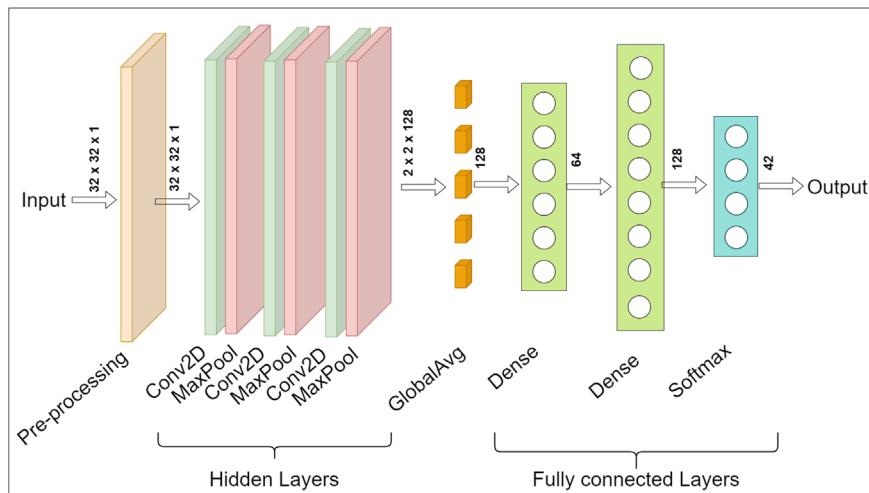
## 2 Methods

Recognition of handwritten characters was performed using two different approaches. In the first approach, conventional CNN layers were used as the primary hidden layers for feature extraction to formulate the baseline model. Whereas in the second approach, pre-trained models were used to replace the CNN layers in the baseline model. EfficientNetB0 [14], ResNet50 [15], MobileNetV2 [16], VGG16

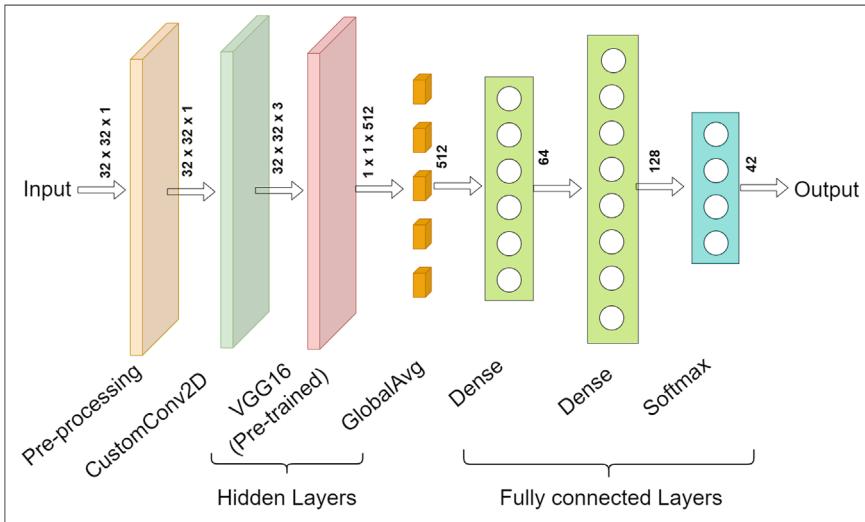
[17] and DenseNet121 [18] were identified as the pre-trained models for the HWCR. Finally, the accuracy and processing time of both the approaches were compared.

Figure 1 shows the architecture of the baseline model. The architecture consists of two major parts: hidden layers and fully connected layers. The depicted architecture begins with an input layer that accepts  $32 \times 32 \times 1$  grayscale images. The hidden layers consist of three convolutional layers (Conv2D) with ReLU activation functions, each of them followed by a max pooling layer to extract and down-sample features. The extracted features were converted into a 128-dimensional vector through global average pooling to avoid overfitting. The fully connected layers process the global averaged vector through dense layers using ReLU activations. Finally the softmax layer with 42 units, produces a probability distribution across 42 output classes. The class corresponding to the highest probability assigned as the output of the given input image.

Apart from the baseline model, pre-trained models were employed to capture the character from small grayscale images. The conventional CNN layers in the baseline model were replaced with the pre-trained models. Figure 2 shows the model architecture in which the conventional CNN layers baseline model is replaced by VGG16. As pre-trained models are often trained on RGB (3 channels) images, the grayscale (1 channel) images are incompatible as input. Hence, a custom CNN layer was introduced to adapt the single-channel images for the pre-trained model. A convolutional 2D layer with number of filters 3 was used as the custom CNN layer. The other layers were fixed as similar to the baseline model. Similar to VGG16, the architecture for other identified pre-trained models were also modified to capture the characters from small grayscale images. Pretrained models such as EfficientNetB0, ResNet50, MobileNetV2, VGG16, and DenseNet121 were the chosen pretrained models. These models were typically trained on large-scale image datasets. The



**Fig. 1** Architecture of the proposed baseline model with conventional CNN layers



**Fig. 2** Architecture of the proposed model with custom CNN with VGG16

primary dataset used for training these models is the ImageNet dataset. ImageNet contains over 14 million annotated images, spanning more than 20,000 categories.

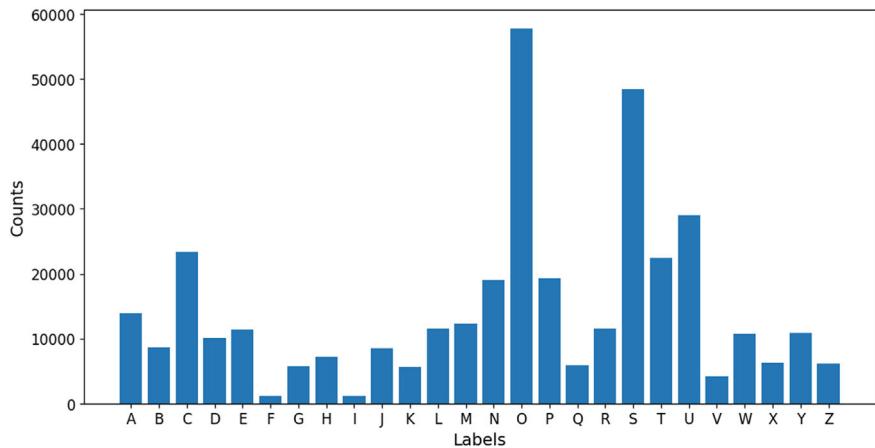
## 2.1 Datasets

Two datasets (A and B) were used in this study. Dataset A contains images of handwritten English capital letters [19]. The dataset consists of 26 classes of 372,450 grayscale images, with  $28 \times 28$  dimension. Figure 3 shows the count of the images in the dataset A with their labels. There was a high imbalance in the dataset.

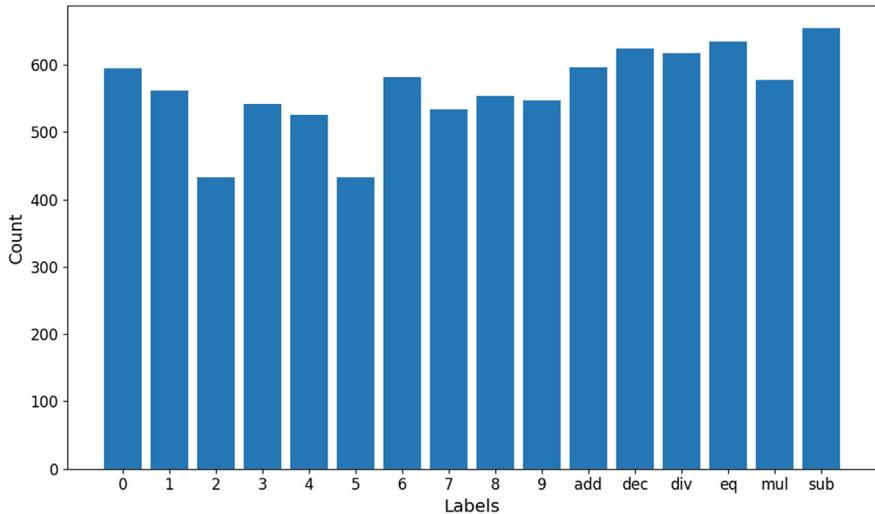
The dataset B consists of 9000 handwritten digits and arithmetic operators [20]. There were a total 16 classes and most images were of resolution  $400 \times 400$  pixels and some were  $155 \times 155$ . Figure 4 shows the count of the images in the dataset B with different labels. Even though the dataset B balanced, the difference in the count between dataset A and B was very high. Total 42 classes were in both datasets.

## 2.2 Preprocessing of Images

Images of dataset A were transformed as  $32 \times 32$  from  $28 \times 28$  by padding. Since the minimum size required for pre-trained models was  $32 \times 32$ . Figure 5 shows a transformed image from dataset A. Images of dataset B were transformed as  $32 \times$



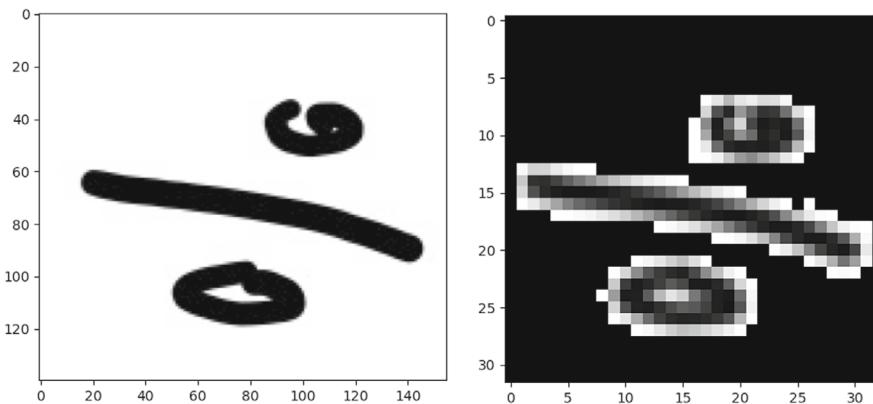
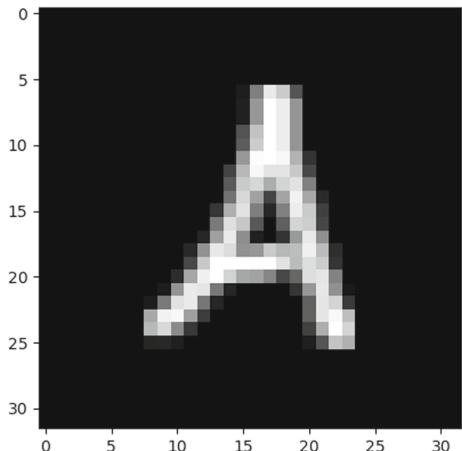
**Fig. 3** Count of the images in the dataset A with different labels



**Fig. 4** Count of the images in the dataset B with different labels

32 from  $400 \times 400$  or  $155 \times 155$  by center crop, and inversion operations were also performed. Figure 6 shows the original and transformed images.

**Fig. 5** A sample transformed image from dataset A

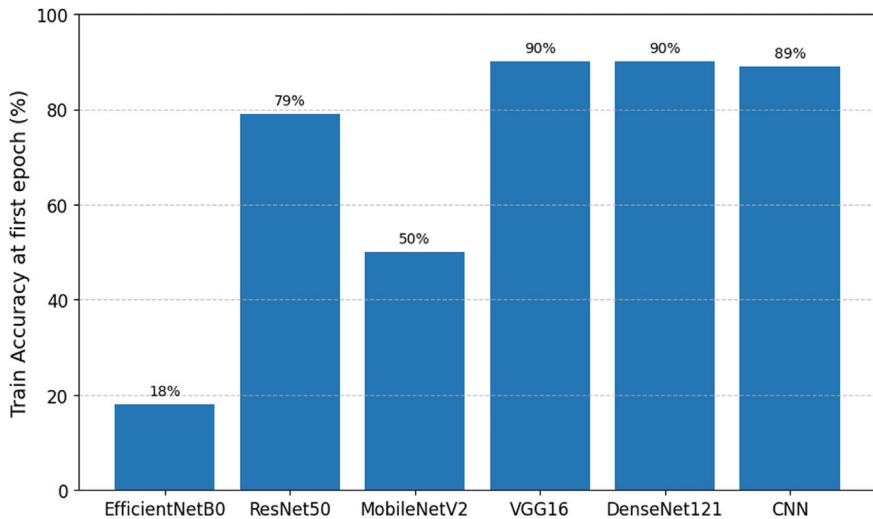


**Fig. 6** A sample original and transformed image from dataset B

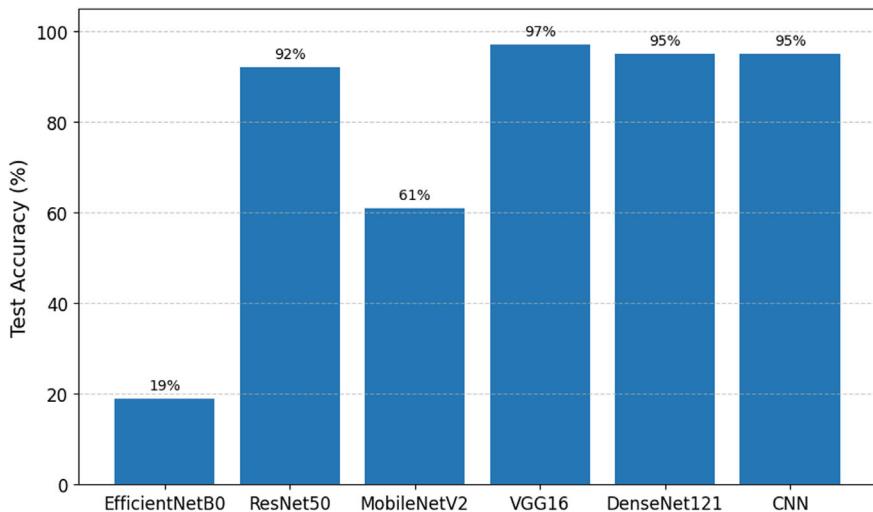
### 3 Results and Discussion

The accuracy of the various models for the first epoch of the training set and the test set, where shown in Figs. 7 and 8 respectively. Figure 7 shows that VGG16, DenseNet121, and CNN were very fast in capturing the pattern in the grayscale images of handwritten characters. With about 90% accuracy, these models reached the first epoch itself. Figure 8 shows that these three models' test accuracy was higher than the others.

Figure 9 shows the time to complete the first epoch for all the models. Conventional CNN and VGG16 took less time than other models. Conventional CNN performance was better than other models. However, VGG16's performance was better than other pre-trained models.

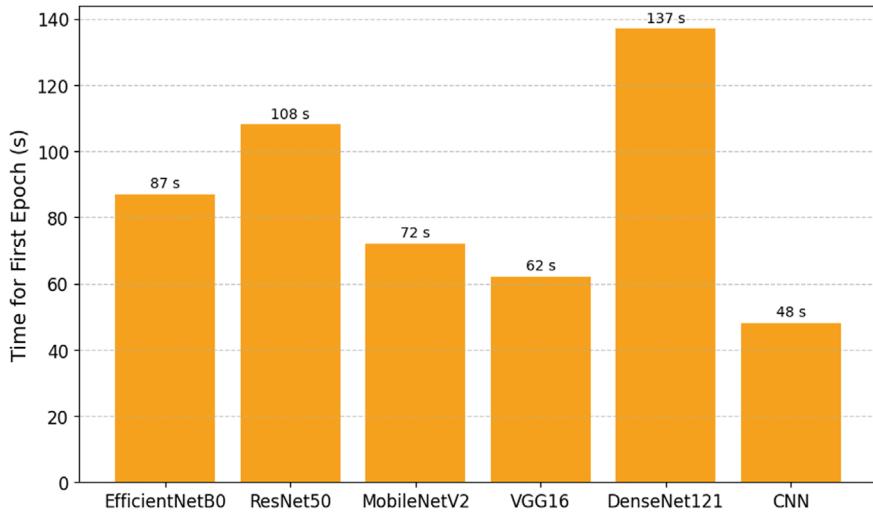


**Fig. 7** Accuracy of different models for the first epoch of training set



**Fig. 8** Accuracy of different models for the test dataset

Confusion matrix is given in Fig. 10. It was found major misclassification happens between the pair of symbols O and D, W and N, N and U, W and U, P and T have a misclassification 131, 95, 72, 73 and 54 respectively.



**Fig. 9** Time taken for the first epoch of different models

True Label	Predicted Label
O - 0	O - 0
O - 0	P - 0
O - 0	Q - 0
O - 0	R - 0
O - 0	S - 0
O - 0	T - 0
O - 0	U - 0
O - 0	V - 0
O - 0	W - 0
O - 0	X - 0
O - 0	Y - 0
O - 0	Z - 0
1 - 0	O - 0
1 - 0	1 - 0
1 - 0	2 - 0
1 - 0	3 - 0
1 - 0	4 - 0
1 - 0	5 - 0
1 - 0	6 - 0
1 - 0	7 - 0
1 - 0	8 - 0
1 - 0	9 - 0
1 - 0	add - 0
1 - 0	div - 0
1 - 0	eq - 0
1 - 0	mul - 0
1 - 0	sub - 0
2 - 0	O - 0
2 - 0	1 - 0
2 - 0	2 - 0
2 - 0	3 - 0
2 - 0	4 - 0
2 - 0	5 - 0
2 - 0	6 - 0
2 - 0	7 - 0
2 - 0	8 - 0
2 - 0	9 - 0
2 - 0	add - 0
2 - 0	div - 0
2 - 0	eq - 0
2 - 0	mul - 0
2 - 0	sub - 0
3 - 0	O - 0
3 - 0	1 - 0
3 - 0	2 - 0
3 - 0	3 - 0
3 - 0	4 - 0
3 - 0	5 - 0
3 - 0	6 - 0
3 - 0	7 - 0
3 - 0	8 - 0
3 - 0	9 - 0
3 - 0	add - 0
3 - 0	div - 0
3 - 0	eq - 0
3 - 0	mul - 0
3 - 0	sub - 0
4 - 0	O - 0
4 - 0	1 - 0
4 - 0	2 - 0
4 - 0	3 - 0
4 - 0	4 - 0
4 - 0	5 - 0
4 - 0	6 - 0
4 - 0	7 - 0
4 - 0	8 - 0
4 - 0	9 - 0
4 - 0	add - 0
4 - 0	div - 0
4 - 0	eq - 0
4 - 0	mul - 0
4 - 0	sub - 0
5 - 0	O - 0
5 - 0	1 - 0
5 - 0	2 - 0
5 - 0	3 - 0
5 - 0	4 - 0
5 - 0	5 - 0
5 - 0	6 - 0
5 - 0	7 - 0
5 - 0	8 - 0
5 - 0	9 - 0
5 - 0	add - 0
5 - 0	div - 0
5 - 0	eq - 0
5 - 0	mul - 0
5 - 0	sub - 0
6 - 0	O - 0
6 - 0	1 - 0
6 - 0	2 - 0
6 - 0	3 - 0
6 - 0	4 - 0
6 - 0	5 - 0
6 - 0	6 - 0
6 - 0	7 - 0
6 - 0	8 - 0
6 - 0	9 - 0
6 - 0	add - 0
6 - 0	div - 0
6 - 0	eq - 0
6 - 0	mul - 0
6 - 0	sub - 0
7 - 0	O - 0
7 - 0	1 - 0
7 - 0	2 - 0
7 - 0	3 - 0
7 - 0	4 - 0
7 - 0	5 - 0
7 - 0	6 - 0
7 - 0	7 - 0
7 - 0	8 - 0
7 - 0	9 - 0
7 - 0	add - 0
7 - 0	div - 0
7 - 0	eq - 0
7 - 0	mul - 0
7 - 0	sub - 0
8 - 0	O - 0
8 - 0	1 - 0
8 - 0	2 - 0
8 - 0	3 - 0
8 - 0	4 - 0
8 - 0	5 - 0
8 - 0	6 - 0
8 - 0	7 - 0
8 - 0	8 - 0
8 - 0	9 - 0
8 - 0	add - 0
8 - 0	div - 0
8 - 0	eq - 0
8 - 0	mul - 0
8 - 0	sub - 0
9 - 0	O - 0
9 - 0	1 - 0
9 - 0	2 - 0
9 - 0	3 - 0
9 - 0	4 - 0
9 - 0	5 - 0
9 - 0	6 - 0
9 - 0	7 - 0
9 - 0	8 - 0
9 - 0	9 - 0
9 - 0	add - 0
9 - 0	div - 0
9 - 0	eq - 0
9 - 0	mul - 0
9 - 0	sub - 0
add - 0	O - 0
add - 0	1 - 0
add - 0	2 - 0
add - 0	3 - 0
add - 0	4 - 0
add - 0	5 - 0
add - 0	6 - 0
add - 0	7 - 0
add - 0	8 - 0
add - 0	9 - 0
add - 0	add - 0
add - 0	div - 0
add - 0	eq - 0
add - 0	mul - 0
add - 0	sub - 0
div - 0	O - 0
div - 0	1 - 0
div - 0	2 - 0
div - 0	3 - 0
div - 0	4 - 0
div - 0	5 - 0
div - 0	6 - 0
div - 0	7 - 0
div - 0	8 - 0
div - 0	9 - 0
div - 0	add - 0
div - 0	div - 0
div - 0	eq - 0
div - 0	mul - 0
div - 0	sub - 0
eq - 0	O - 0
eq - 0	1 - 0
eq - 0	2 - 0
eq - 0	3 - 0
eq - 0	4 - 0
eq - 0	5 - 0
eq - 0	6 - 0
eq - 0	7 - 0
eq - 0	8 - 0
eq - 0	9 - 0
eq - 0	add - 0
eq - 0	div - 0
eq - 0	mul - 0
eq - 0	sub - 0
mul - 0	O - 0
mul - 0	1 - 0
mul - 0	2 - 0
mul - 0	3 - 0
mul - 0	4 - 0
mul - 0	5 - 0
mul - 0	6 - 0
mul - 0	7 - 0
mul - 0	8 - 0
mul - 0	9 - 0
mul - 0	add - 0
mul - 0	div - 0
mul - 0	eq - 0
mul - 0	sub - 0
sub - 0	O - 0
sub - 0	1 - 0
sub - 0	2 - 0
sub - 0	3 - 0
sub - 0	4 - 0
sub - 0	5 - 0
sub - 0	6 - 0
sub - 0	7 - 0
sub - 0	8 - 0
sub - 0	9 - 0
sub - 0	add - 0
sub - 0	div - 0
sub - 0	eq - 0
sub - 0	mul - 0

**Fig. 10** Confusion matrix

## 4 Conclusion

The study discusses adjustment of pre-trained models in identifying handwritten characters and mathematical symbols from grayscale images with a single channel. A customized CNN layer is introduced to convert grayscale images into three channels, allowing the utilisation of different pre-trained models including EfficientNetB0, ResNet50, MobileNetV2, VGG16, and DenseNet121. Out of these models, VGG16, DenseNet121, and a traditional CNN demonstrated exceptional performance, attaining an accuracy of roughly 90% in the initial epoch and around 96% accuracy in the testing phase. The standard CNN and VGG16 models exhibited superior accuracy and faster training times, making them highly efficient for character recognition in small grayscale images. The results emphasise the efficacy of pre-trained models in improving HWCR systems, offering a strong method for precise and efficient recognition of handwritten characters across many categories, including mathematical symbols. The misclassification of closely resembling characters can be reduced further by processing a sequence of characters as word with Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks which would capture the dependencies and contextual information. The limitation of the study was imbalance datasets with a smaller number of images. For future research, with a high number of balanced datasets, the above said models could be developed to compare their performance.

**Author Contribution** **Manibharathi D and Sanjeev Kumar:** Conceptualization, Methodology, Software, Writing-Original draft preparation. **Vasanthanayaki C:** Supervision, Reviewing.

## References

1. Zhou Y, Zuo S, Yang Z, He J, Shi J, Zhang R (2023) A review of document image enhancement based on document degradation problem. *Appl Sci* 13(13):7855
2. Li Y, Yang Q, Chen Q, Hu B, Wang X, Ding Y, Ma L (2023) Fast and robust online handwritten Chinese character recognition with deep spatial and contextual information fusion network. *IEEE Trans Multimed* 25:2140–2152
3. Zhang W, Sun C, Gao Y (2023) Image intensity variation information for interest point detection. *IEEE Trans Pattern Anal Mach Intell* 45:9883–9894
4. Bhati GS, Garg AR (2021) CNN-based handwritten mathematical symbol recognition model. *Cyber Intell Inf Retr* 407–416
5. Ghosh T, Abedin MHZ, Al Banna H et al (2021) Performance analysis of state of the art convolutional neural network architectures in Bangla handwritten character recognition. *Pattern Recognit Image Anal* 31:60–71
6. Altwaijry N, Al-Turaiki I (2020) Arabic handwriting recognition system using convolutional neural network. *Neural Comput Appl* 33(7):2249–2261
7. Singh S, Sharma A, Chauhan VK (2021) Online handwritten Gurmukhi word recognition using fine-tuned deep convolutional neural network on offline features. *Mach Learn Appl* 5:100037
8. Shetty A, Sharma S (2024) Ensemble deep learning model for optical character recognition. *Multimed Tools Appl* 83:11411–11431
9. Mushtaq F, Misgar MM, Kumar M (2021) UrduDeepNet: offline handwritten Urdu character recognition using deep neural network. *Neural Comput & Appl* 33:15229–15252

10. Benissa A, Bahri A, El Allaoui A, Bourass Y (2022) Character recognition using pre-trained models and performance variants based on datasets size: a survey. In: ITM web of conferences, vol 43. pp 01008
11. Bhati GS, Garg AR (2020) Handwritten Devanagari character recognition using CNN with transfer learning. Congr Intell Syst. Springer, Singapore
12. Chandure S, Inamdar V (2023) Handwritten MODI character recognition using transfer learning with discriminant feature analysis. IETE J Res 69(5):2584–2594
13. Albattah W, Albahli S (2022) Intelligent Arabic handwriting recognition using different standalone and hybrid CNN architectures. Appl Sci
14. Tan M, Le QV (2019) EfficientNet: rethinking model scaling for convolutional neural networks. In: Proceedings of the 36th international conference on machine learning (ICML), vol 97. pp 6105–6114
15. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). pp 770–778
16. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC (2018) MobileNetV2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). pp 4510–4520
17. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: Proceedings of the international conference on learning representations (ICLR)
18. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). pp 4700–4708
19. [https://iaexpert.academy/arquivos/alfabeto\\_A-Z.zip](https://iaexpert.academy/arquivos/alfabeto_A-Z.zip). Accessed 13 Apr 2024
20. <https://www.kaggle.com/datasets/sagyamthapa/handwritten-math-symbols/data>. Accessed 13 Apr 2024

# Forecasting Heart Disease Using Deep BI-DI Neural Networks



R. Bhuvanya , T. Kujani , and P. Matheswaran

**Abstract** Globally, heart disease continues to claim countless lives, remaining a top killer, highlighting the need for accurate prediction models to aid in early diagnosis and intervention. To enhance heart attack prediction and improve patient outcomes, this study investigates the efficacy of a deep Bi-Directional Recurrent Neural Network architecture. This novel approach leverages stacked Long Short-Term Memory (LSTM) units to capture temporal dependencies within medical data, potentially surpassing the accuracy of conventional Machine Learning (ML) algorithms in predicting heart attacks. The results highlight that the twin Long Short-Term Memory model achieved an exceptional accuracy of 98% in predicting heart attacks, surpassing other conventional ML algorithms. The results highlight the significance of utilizing sophisticated machine learning techniques in the medical field to enhance patient results and decrease the prevalence of cardiovascular illnesses.

**Keywords** Heart attack prediction · Machine learning · Model evaluation · Risk factors · Stacked long short-term memory

## 1 Introduction

An artery to the heart being blocked is what causes a heart attack. Plaques are formed as fatty deposits and cholesterol build up over time in the arteries leading to the heart. Heart attacks and strokes belong to cardiovascular disorders. Heart

---

R. Bhuvanya ()

Sri Ramachandra Faculty of Engineering and Technology, Sri Ramachandra Institute of Higher Education and Research, Chennai, Tamil Nadu 600016, India

e-mail: [bhuvanyaraghunathan@gmail.com](mailto:bhuvanyaraghunathan@gmail.com)

T. Kujani

VelTech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu 600062, India

P. Matheswaran

K. Ramakrishnan College of Technology, Samayapuram, Tiruchirappalli, Tamil Nadu 621112, India

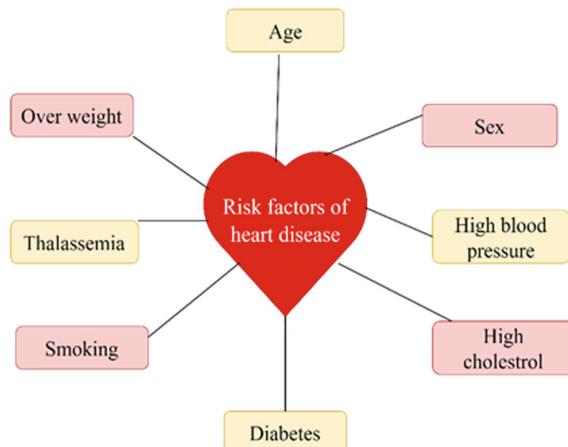
attacks also known as myocardial infarctions cover more than any other disease worldwide [1]. Common risk factors for a heart attack are advanced age (with risk escalating as age rises), diabetes, being overweight, physical inactivity, smoking, high cholesterol levels, and diabetes [2]. Other indications of a heart attack encompass chest discomforts arm pain episodes of nausea and vomiting. Figure 1 displays risk factors for cardiac diseases. According to estimates by World Health Organization, (WHO) approximately seventy-five percentage of premature heart disease can be prevented and lowering risk factors might reduce today's increasing burden of these diseases on patients and health care providers [3]. Detecting symptoms early enough as well as intervening is however essential for improved patient outcomes but conventional approaches often rely on late-stage symptoms or invasive tests. This is where machine learning (ML) comes into play providing a non-invasive way to predict and prevent coronary attacks. This means that their accuracy and predictive power can continuously improve over time, leading to better outcomes for patients.

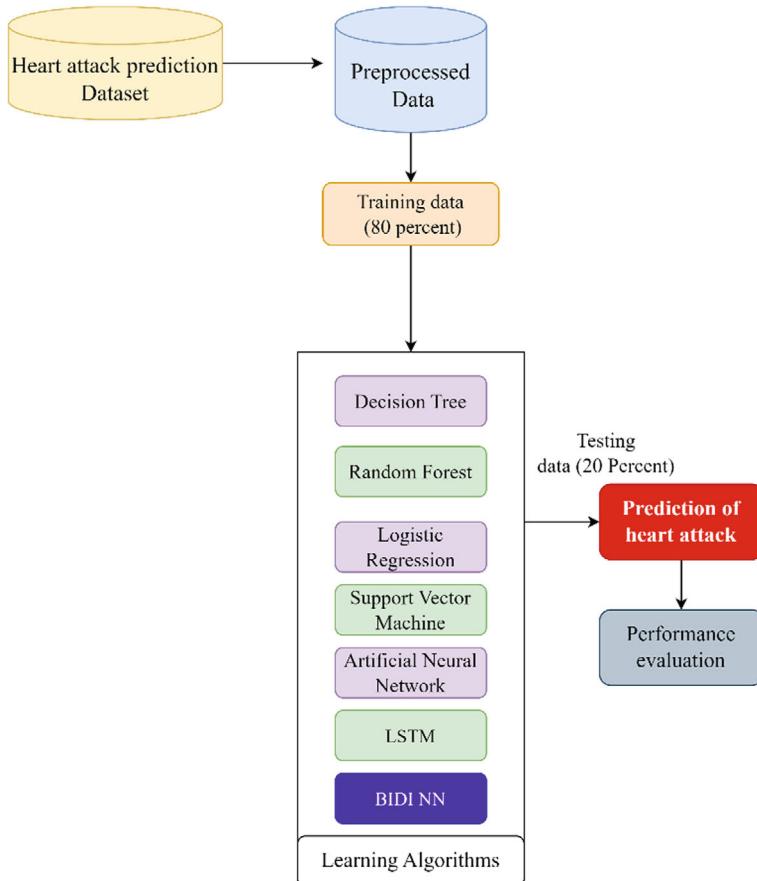
Previous research has explored applying machine learning and deep learning for cardiac disease prediction [4–6]. ML models may examine population-level data in order to identify trends, patterns, and risk factors associated with heart attacks. By reducing the risk of heart disease and promoting cardiovascular well-being, these statistics can subsequently guide improvements to public health policies and interventions. Figure 2 presents an overall architectural representation of the proposed work.

The subject of machine learning-based cardiac disease prediction gains significant advancements with this article. This paper presents significant contributions to heart disease prediction in the field of machine learning.

**Comprehensive Benchmarking:** It provides a comparative analysis of established machine learning models like Decision Trees, Random Forest, Logistic Regression, Support Vector Machine, and Artificial Neural Networks. This comparison establishes a baseline performance for heart disease prediction.

**Fig. 1** Risk factors of heart disease





**Fig. 2** Architecture of the proposed solution

**Introduction of a Novel Architecture:** Beyond the traditional models, the paper proposes a deep BI-DIrectional Neural Network (BI-DI\_NN) specifically designed for heart disease prediction. This new architecture demonstrates superior performance compared to existing methods.

**Optimizing Network Depth:** The study investigates the impact of network depth on prediction accuracy. By evaluating twin NN (2 layers), shallow NN (3 layers), Mid-depth NN (4 layers), and Deep NN (5 layers), it identifies the most suitable network architecture for this specific task. This finding can be valuable for researchers seeking to optimize neural network design for heart disease prediction.

Four sections make up the structure of the paper. It is important to forecast cardiac disease, as the first part emphasizes. An in-depth analysis of the current research techniques for heart disease prediction is given in the second section. Our proposed methodology is thoroughly explained in depth in the third section. In the end, the performance evaluation of our suggested method is shown in the fourth part.

## 2 Literature Survey

Researchers have been actively developing machine learning methods to improve heart disease prediction using patient medical records. Here's a summary of some key findings, showcasing the steady improvement in accuracy and the variety of approaches taken: Early works like Buchan et al. laid the foundation by developing new methods to assess heart disease risk from medical records. Their approach involved creating a unique feature extraction technique based on an ontology [7]. More recent studies like Al-Tashi et al. and Ashraf et al. have achieved impressive accuracy rates exceeding 90% using Support Vector Machines and deep neural networks respectively. Al-Tashi et al. implemented a wrapper feature selection method based on grey wolf optimization, while Ashraf et al. proposed a deep neural network with multiple hidden layers to improve feature quality [4, 8]. Feature selection techniques have been a prominent area of research. Mohan et al. specifically focused on identifying key features for better heart disease prediction accuracy. They achieved an accuracy of 88.7% using a model that combines feature extraction and selection methods. Similarly, Gokulnath et al. aimed to optimize feature selection using a Genetic Algorithm with Support Vector Machines, achieving an accuracy of 88.34% [6, 9]. Classification algorithms have also been explored. Karadeniz et al. presented two distinct classifiers, finding that the Shrunk Covariance Classifier outperformed others on a specific dataset [10]. Patro et al. took a different approach, combining various techniques for feature selection and classification. Their model using a Bayesian-optimized Support Vector Machine achieved an accuracy of 93.3% [11]. Deep learning approaches are pushing the boundaries of accuracy. Deepika et al. proposed a method combining an optimized unsupervised feature selection technique with a Multi-Layer Perceptron, achieving an accuracy of 94.28%. Barfungpa et al. went a step further, achieving an accuracy of 99.57% using a deep learning model that integrates data mining techniques for disease prediction. Their model employs multiple steps including data preprocessing, feature selection, and a deep neural network architecture [12, 13]. Data quality is also a crucial factor. Namasudra et al. proposed a Data Trust Model to improve the quality of time series data used for predictions. They further introduced an enhanced Neural Network model that incorporates this approach [14]. These studies highlight the potential of machine learning for heart disease prediction, with continuous advancements in achieving higher accuracy. Deep learning models are producing remarkable results, but alternative approaches can also be successful. The selection of the appropriate method may vary based on factors such as the dataset in use and the available computational resources.

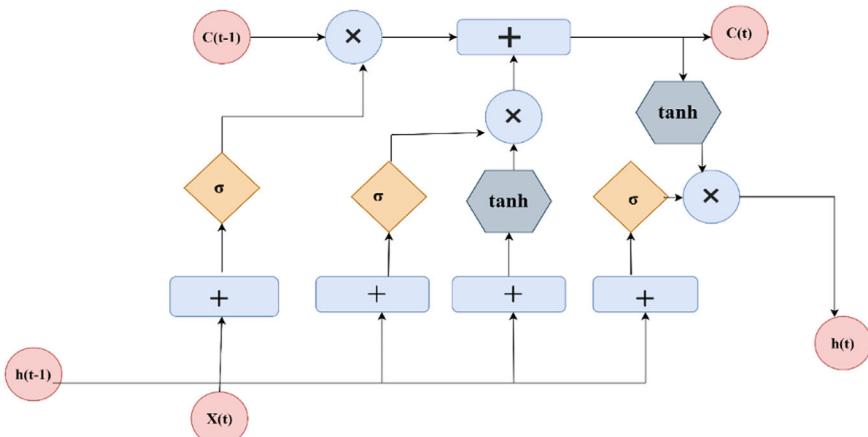
### 3 Proposed Method

Traditional methods of predicting heart attacks often rely on simpler statistical models and manual assessments by healthcare professionals. Machine learning algorithms, particularly those using complex models like neural networks, can uncover patterns and correlations that might be missed by human analysis. These models can process and analyze data more quickly and accurately, leading to better risk stratification and personalized care. This paper proposes a Bi-Directional LSTM for better prediction of heart disease.

#### 3.1 BI-DIrectional LSTM

Long Short-Term Memory (LSTM) networks tackle the issue of the vanishing gradient, which is a challenge for conventional Recurrent Neural Networks (RNNs) [15]. These LSTM models are capable of being trained on fresh data, allowing them to adjust to new patterns and advancements in medical understanding. This leads to more reliable and current models, which in turn enhance the predictive power over time. The essential parts of LSTM are the input, short-term memory, and long-term memory. The critical elements of LSTM are the input gate, forget gate, cell state, and output gate. Figure 3 clearly illustrates the structure of LSTM.

The input gate controls the entry of new information. It employs a sigmoid activation function ( $\sigma$ ) to assess the importance of this new information.  $\lambda$  denotes the hidden state of the LSTM cell at the time of the previous step ( $t - 1$ ), and  $\alpha$  represents the current input.



**Fig. 3** Illustration of LSTM

$$\text{Input}(t) = \sigma(W * [\lambda, \alpha] + bi) \quad (1)$$

where  $W$  is the weight matrix of input gate, and  $bi$  is the Input gate bias vector.

The forget gate selectively remembers information from the prior cell state. It employs a sigmoid activation function ( $\sigma$ ) to output values between 0 and 1, where 0 indicates complete forgetting and 1 signifies full retention. Mathematically

$$\text{forget}(t) = \sigma(Wf * [\lambda, \alpha] + bf) \quad (2)$$

where,  $Wf$  is the weight matrix of forget gate,  $bf$  is the forget gate bias vector and  $\sigma$  is the Sigmoid activation function (outputs values between Yes or No).

The cell state  $\text{cell}(t)$  is updated by merging the forgotten information with the newly allowed information.

$$\text{cell}(t) = \tanh(Wc * [\lambda, \alpha] + bc) \quad (3)$$

where,  $Wc$  is the weight matrix of cell state,  $bc$  is the candidate cell state bias vector and  $\tanh$  is the tangent activation function.

The output gate, denoted as  $o(t)$ , determines the specific information from the current cell state  $c(t)$  that will be incorporated into the hidden state  $h(t)$ , thus serving as the LSTM cell's output at this particular time step.

$$o(t) = \sigma(Wo * [\lambda, \alpha] + bo) \quad (4)$$

The above steps are repeated for each time step in the sequence, allowing the LSTM to learn long-term dependencies within the data.

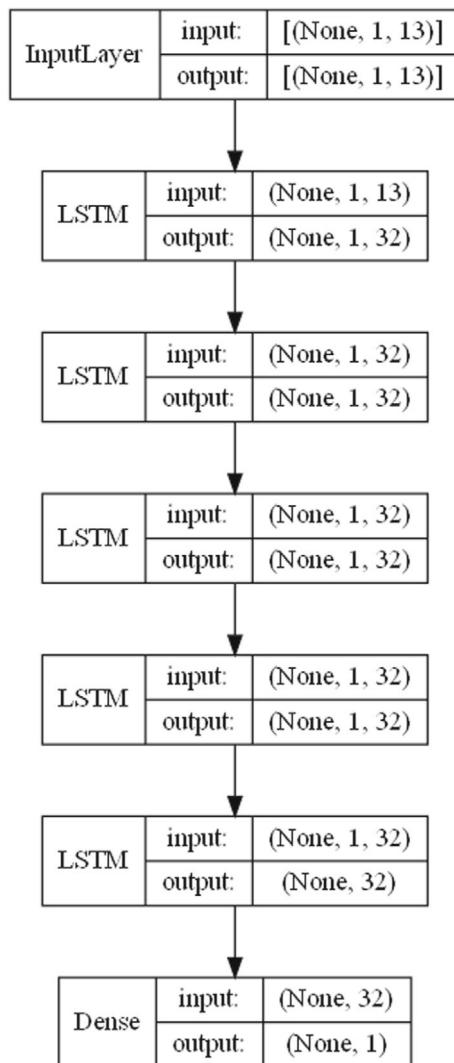
where  $h(t)$  is the output of the current block,  $\tanh$  is the hyperbolic tangent,  $\times$  stands for element-wise multiplication,  $+$  is the element-wise summation, and  $\sigma$  is the sigmoid activation function.  $X(t)$  is the input vector,  $C(t - 1)$  is the memory from the previous block.

The proposed BI-DIrectional Neural Network (BIDI\\_NN) is a variant of recurrent neural networks (RNNs) designed to improve the performance of the LSTM. In this architecture, multiple layers of densely connected recurrent units are stacked on top of each other. Unlike traditional RNN architectures, the bi-directional neural network has dense connections between adjacent layers, allowing it to maintain and update internal states across time steps. This allows the network to identify the order of dependencies within the data it receives. The network can acquire layered models of data sequences, capturing both immediate and distant relationships, by employing deep stacking. Because every neuron in a layer receives information from every other neuron in the layer before it, the input sequence creates a completely connected architecture that enables each layer to extract more abstract properties.

In this work, the models with twin NN, shallow NN, mid-depth NN, and deep NN are investigated to establish the ideal number of layers for the BI-DI\\_NN. The

optimum model for heart attack prediction is discovered by comparing the performance. The model architecture of deep NN is portrayed in Fig. 4. The algorithm for stacked LSTM is depicted in Table 1.

**Fig. 4** Model architecture of Deep NN



**Table 1** Procedure for stacked LSTM

Procedure Stacked_LSTM	
<b>Input:</b> Clinical and demographic features of individuals	
<b>Output:</b> prediction of heart disease- Binary classification	
1	Initialize a sequential model
2	for i = 1 to num_layers do:
3	if i == 1 then:
4	Add LSTM layer with hidden_units[i] units and input_shape=(sequence_length, input_dim) to the model
5	else if i == num_layers then:
6	Add LSTM layer with hidden_units[i] units to the model
7	else:
8	Add LSTM layer with hidden_units[i] units and return_sequences=True to the model
9	end for

## 4 Performance Analysis

### 4.1 Datasets and Visualization

Data has been collected from Kaggle [<https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>] which is composed of thirteen columns that represent the features of the data set. The visualization graph of the obtained dataset based on predicting the heart attack is illustrated below. The probability of heart attack prediction based on age, cholesterol level, heart rate, resting\_ecg, thallium, resting blood pressure are depicted in Figs. 5, 6, 7, 8, 9 and 10 respectively.

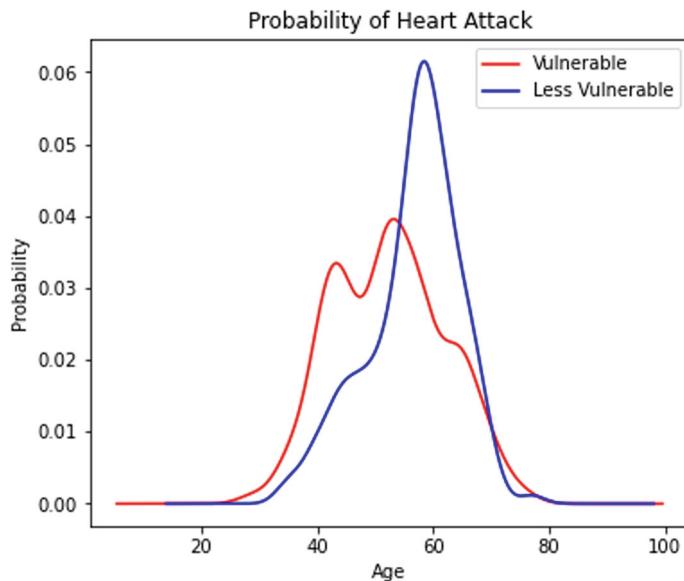
### 4.2 Evaluation Metrics

The following metrics are utilized to assess the effectiveness of the implemented models.

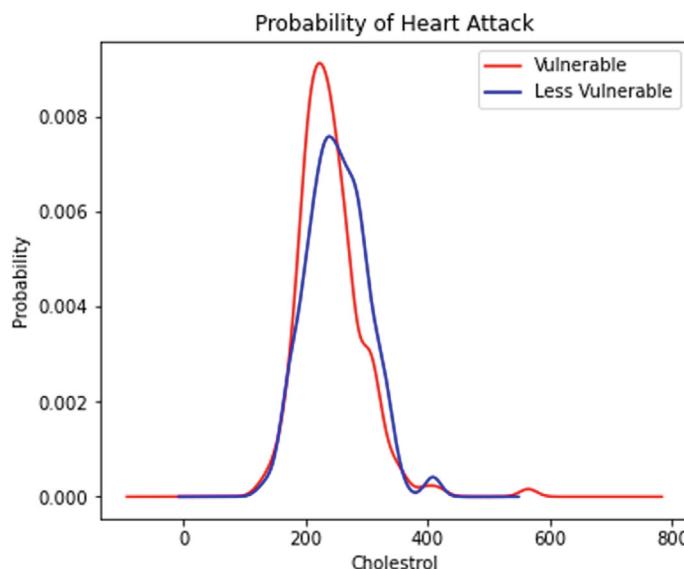
**Accuracy:** It is a metric that shows the total percentage of accurate predictions made by the model.

$$\text{Accuracy} = (\text{T.PE} + \text{T.NE}) / (\text{T.PE} + \text{T.NE} + \text{F.PE} + \text{F.NE}) \quad (5)$$

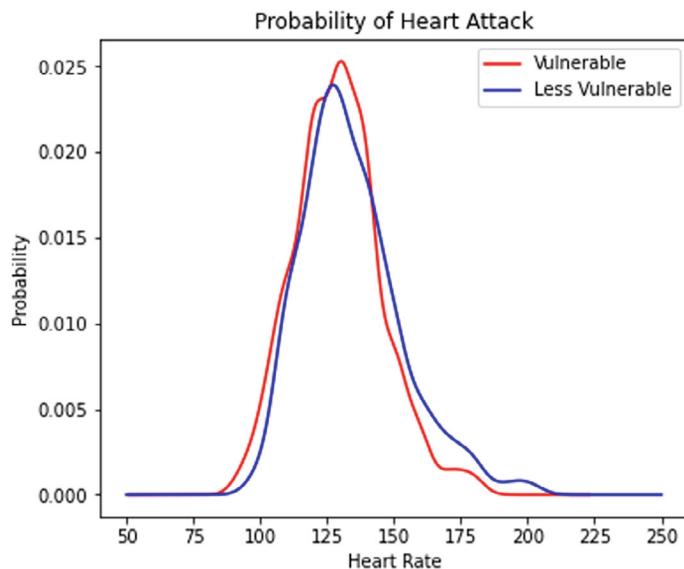
**Precision:** This metric emphasizes accurate positive predictions, showing the ratio of true positive predictions to all instances predicted as positive by the model.



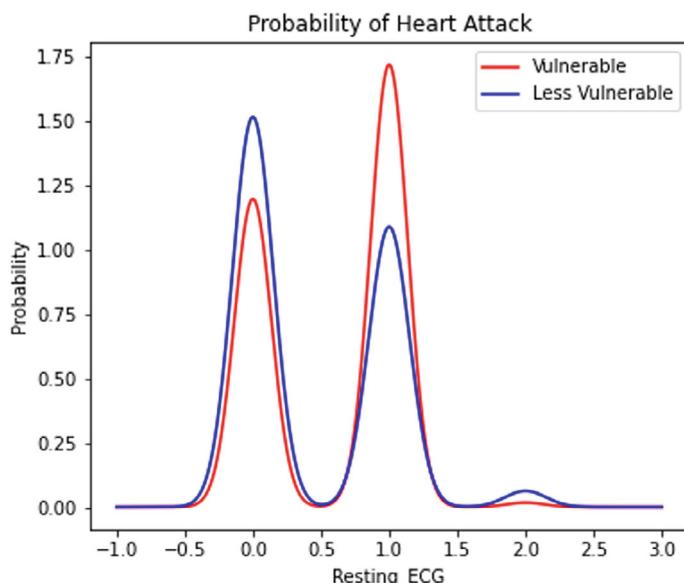
**Fig. 5** Probability of attack based on age



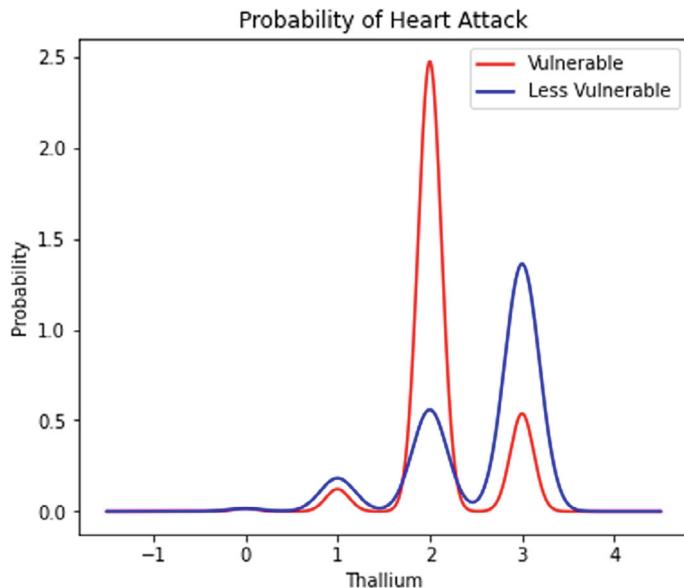
**Fig. 6** Probability of attack based on cholesterol



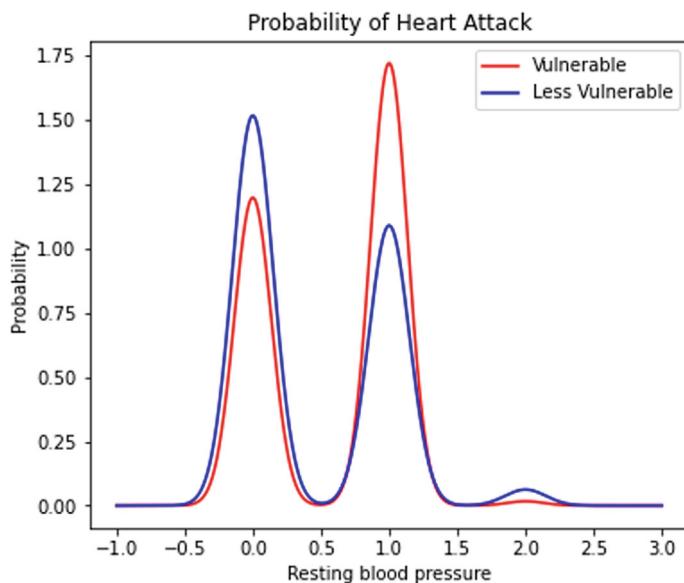
**Fig. 7** Probability of attack based on minimum heart rate



**Fig. 8** Probability of attack based on maximum heart rate



**Fig. 9** Probability of attack based on thallium



**Fig. 10** Probability of attack based on resting blood pressure

$$\text{Precision} = \text{T.PE}/(\text{T.PE} + \text{F.PE}) \quad (6)$$

**Recall:** This measurement highlights the model's capacity to accurately detect real positive instances.

$$\text{Recall} = \text{T.PE}/(\text{T.PE} + \text{F.NE}) \quad (7)$$

**F1 Score:** Mean of precision and recall yields a unified score that considers both precision and recall equally.

$$\text{F1 Score} = 2 * (\text{PN} * \text{RL})/(\text{PN} + \text{RL}) \quad (8)$$

where,

T.PE—True Positive, T.NE—True Negative, F.PE—False Positive and F.NE—False Negative, PN—precision and RL—Recall respectively.

### 4.3 Results and Discussion

The LSTM architecture effectively captured the sequential nature of heart disease data, particularly the evolution of risk factors over time, leading to superior prediction accuracy compared to static models. The model with Twin NN achieved the best balance between capturing long-term dependencies and avoiding overfitting. Increasing the number of layers resulted in a slight performance decrease, suggesting potential overfitting issues. Tables 2, 3 and 4 highlight the results of precision, recall, and F1 score for the adopted machine learning models. Figure 11 depicts the graphical representation of accuracy.

**Table 2** Results of precision

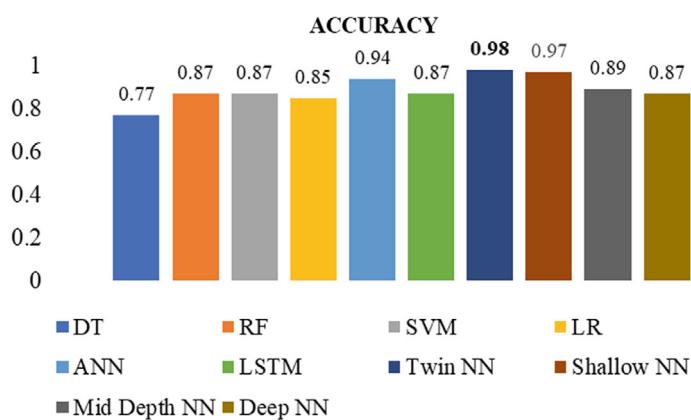
Techniques applied	Less probability heart attack	High probability heart attack
Dec_Tree	0.71	0.85
Ran_F	0.84	0.90
SupportVM	0.82	0.93
Log_R	0.86	0.85
ANN	0.93	0.95
LSTM	0.86	0.87
Twin NN	<b>0.97</b>	<b>1.00</b>
Shallow NN	0.95	0.99
Mid Depth NN	0.92	0.86
Deep NN	0.83	0.92

**Table 3** Results of recall

Techniques applied	Less probability heart attack	High probability heart attack
Dec_Tree	0.86	0.69
Ran_F	0.90	0.84
SupportVM	0.93	0.81
Log_R	0.83	0.88
ANN	0.95	0.93
LSTM	0.87	0.87
<b>Twin NN</b>	<b>1.00</b>	<b>0.97</b>
Shallow NN	0.99	0.95
Mid Depth NN	0.85	0.93
Deep NN	0.92	0.87

**Table 4** Results of F1

Techniques applied	Less probability heart attack	High probability heart attack
Dec_Tree	0.87	0.87
Ran_F	0.87	0.87
SupportVM	0.87	0.87
Log_R	0.84	0.86
ANN	0.94	0.94
LSTM	0.87	0.87
<b>Twin NN</b>	<b>0.98</b>	<b>0.99</b>
Shallow NN	0.97	0.97
Mid Depth NN	0.88	0.90
Deep NN	0.87	0.86

**Fig. 11** Accuracy level of heart attack prediction

The study tested a deep Bi-Directional Neural Network with different layers. Surpassing all baseline machine learning models, the dual neural network achieved an impressive accuracy of 98%. This shows that the model is able to accurately identify patterns in the data and make accurate predictions. This finding suggests that the utilization of neural networks, specifically the twin NN architecture, holds promise for accurate and reliable heart attack prediction.

## 5 Conclusion and Future Enhancements

In conclusion, the work evaluated multiple machine learning classifiers for predicting heart attacks, and among all the methods explored, the twin neural network (NN) model demonstrated the highest accuracy in predicting heart attacks. The excellent performance of Twin NN highlights the importance of using advanced neural network architectures in medical predictive modeling tasks. It must be understood that the success of a neural network model does not depend only on increasing the number of layers. Instead, the optimal number of layers must be carefully chosen based on the characteristics and complexity of the dataset. This highlights the importance of selecting an appropriate architecture tailored to the specific requirements of the data at hand, as opposed to blindly increasing the network's depth.

While the Bi-Directional NN architecture achieved high accuracy, investigating the effectiveness of more sophisticated deep learning architectures is warranted. Convolutional Neural Networks (CNNs) have shown promise in processing image data, and transformers have demonstrated success with sequential data. Exploring these architectures in the future for heart disease prediction, potentially in conjunction with the Bi-Directional NN, could lead to further performance improvements.

## References

1. Vaduganathan M, Mensah G, Turco J et al (2022) The global burden of cardiovascular diseases and risk: a compass for future health. *J Am Coll Cardiol* 80(25):2361–2371. <https://doi.org/10.1016/j.jacc.2022.11.005>
2. WHO (2016) The challenge of cardiovascular disease—quick statistics. <http://www.euro.who.int/en/health-topics/noncommunicable-diseases/cardiovascular-diseases/data-and-statistics>. Accessed 10 October 2016
3. Benjamin Fredrick David H, Antony Belcy S (2018) Heart disease prediction using data mining techniques. *ICTACT J Soft Comput* 9(1)
4. Ashraf M, Rizvi MA, Sharma H (2019) Improved heart disease prediction using deep neural network. *Asian J Comput Sci Technol* 8(2):49–54
5. Gárate-Escamila AK, El Hassani AH, Andrès E (2020) Classification models for heart disease prediction using feature selection and PCA. *Inform Med Unlocked* 19:100330
6. Gokulnath CB, Shanthalrajah SP (2019) An optimized feature selection based on genetic approach and support vector machine for heart disease. *Clust Comput* 22:14777–14787
7. Buchan K, Filannino M, Uzuner Ö (2017) Automatic prediction of coronary artery disease from clinical narratives. *J Biomed Inform* 72:23–32

8. Al-Tashi Q, Rais H, Jadid S (2018) Feature selection method based on grey wolf optimization for coronary artery disease classification. In: International conference of reliable information and communication technology, pp 257–266
9. Mohan SC, Thirumalai C, Srivastava G (2019) Effective heart disease prediction using hybrid machine learning techniques. IEEE Access 7:81542–81554
10. Karadeniz T, Tokdemir G, Maraş HH (2021) Ensemble methods for heart disease prediction. N Gener Comput 3(3):569–581
11. Patro SP, Nayak GS, Padhy N (2021) Heart disease prediction by using novel optimization algorithm: a supervised learning prospective. Inform Med Unlocked 26:100696
12. Deepika D, Balaji N (2022) Effective heart disease prediction using novel MLP-EBMDA approach. Biomed Signal Process Control 72:103318. Part B
13. Barfungpa SP, Sarma HKD, Samantaray L (2023) An intelligent heart disease prediction system using hybrid deep dense Aquila network. Biomed Signal Process Control 84:104742
14. Namasudra S, Dhamodharavadhani S, Rathipriya R, Crespo RG, Moparthi NR (2024) Enhanced neural network-based univariate time-series forecasting model for big data. Big Data 12(2):83–99
15. Liang H, Yu Y, Jiang L, Xie Z (2019) SEML: a semantic LSTM model for software defect prediction. IEEE Access 7:83812–83824
16. Gupta C, Saha A, Reddy NS, Acharya UD (2022) Cardiac disease prediction using supervised machine learning techniques. In: Journal of physics: conference series, vol 2161, no 1. IOP Publishing, p 012013
17. Bhuvanya R, Kavitha M (2023) A real-time e-commerce accessories recommender system by coupling deep learning and histogram features. J Intell Fuzzy Syst 45(1):1179–1193. <https://doi.org/10.3233/JIFS-223754>
18. Das SK, Namasudra S, Sangaiah AK (2024) HCNet: hybrid convolution neural network for automatic identification of ischaemia in diabetic foot ulcer wounds. Multimedia Syst 30:36. <https://doi.org/10.1007/s00530-023-01241-4>
19. Bhatia M, Bhatia S, Hooda M et al (2022) Analyzing and classifying MRI images using robust mathematical modeling. Multimed Tools Appl 81:37519–37540. <https://doi.org/10.1007/s11042-022-13505-8>

# Emo-Tune: Harnessing Emotion-Based Music for Patient Wellness



T. Sudha, N. Jothy, V. Bharathi, and Santhosh Jayagopalan

**Abstract** Music has been a popular way to convey and comprehend human emotions, and it is a powerful medium of expression. To enhance emotional well-being, this paper involves identifying an individual's emotional state and playing music that matches their current feelings. Existing models of emotion-based music recommendation are based on various algorithms like Principal Component Analysis (PCA), Support Vector Machines (SVM), Artificial Neural Network (ANN), etc. However, those models have not yielded optimal results and are still being pursued. This paper aims to improve these limitations by proposing a simple, user-friendly Machine Learning (ML) system that can determine human emotional state. This research endeavour is centered around training an ML model with extensive datasets of facial emotions as input training data from the user. This system uses physiological data from the user's face to identify emotional patterns and affinities and suggests playlists, music tracks, and albums that match the recognized emotion. These models intend to capture the subtleties of various emotional states, such as joy, sadness, neutrality, rock, and surprise. This paper is a significant advance in the use of machine learning and music, presenting an innovative method for deeply linking music to human emotions.

**Keywords** Machine learning · Convolutional neural network · Human emotions

---

T. Sudha (✉) · V. Bharathi

Sri Manakula Vinayagar Engineering College, Puducherry, India  
e-mail: [sudha.ice@smvec.ac.in](mailto:sudha.ice@smvec.ac.in)

V. Bharathi

e-mail: [bharathiv@smvec.ac.in](mailto:bharathiv@smvec.ac.in)

N. Jothy

SRM Valliammai Engineering College, Kattangulathur, Tamil Nadu, India  
e-mail: [jothyn.ece@srmvalliammai.ac.in](mailto:jothyn.ece@srmvalliammai.ac.in)

S. Jayagopalan

British Applied College, Umm Al Quwain, UAE  
e-mail: [santhosh.j@acuq.ae](mailto:santhosh.j@acuq.ae)

## 1 Introduction

Music therapy is a complementary medicine modality that utilizes music to improve the well-being of individuals in different ways, such as social, emotional, cognitive, and physical. This therapeutic approach encompasses a spectrum of activities, ranging from vocalization, instrumental performance, and listening to various forms of music, including natural sounds like birdsongs or waterfalls, and engaging in religious chants. Active music therapy involves patient participation in playing instruments or singing, while passive music therapy only involves patients listening to music. An initial investigation was performed on students who were experiencing exam anxiety, evaluating its effects on both positive and negative outcomes.

The findings indicate that music therapy, particularly when using natural sounds, encourages a positive mindset and reduces stress after challenging academic experiences [1]. The potential for music therapy to enhance mental health, emotional well-being, and cognitive functions has been widely acknowledged. The efficiency and personalization of music therapy can be improved by incorporating technology, such as CNNs, to analyze facial expressions.

The paper, ‘Emo-tune: Harnessing Emotion-Based Music for Patient Well-being’ is a significant advance in the use of innovative technology and helpful treatments. Smart technology is utilized to detect emotions on people’s faces and then recommends music that is just right for each individual. The special part is that it can change and choose the music therapy based on how each person is feeling, which seems like a good way to make treatments work even better. Providing healthcare that caters to individual needs while ensuring privacy and security is a beacon of hope. It plays a significant role in the ongoing transformation of healthcare practices, providing a more personalized and emotionally resonant way to address patient wellness. ‘Emo-tune: Harnessing Emotion-Based Music for Patient Wellness’ has promising potential and opens up many possibilities for improvements and wider applications. Systems that suggest music based on your emotions are looking very promising in the future. These systems are improving their ability to comprehend feelings, even when there are subtle differences and how emotions are expressed in different cultures. The idea of using wearable technology like smartwatches is exciting because it allows the system to quickly adjust the music it suggests based on your current emotions. Adding additional information, such as your heart rate or brainwaves, could enhance the accuracy of the music suggestions. They are considering integrating it with other smart helpers on your devices, so it’s not just about music—it could also provide other types of support. They comprehend that there are significant considerations to be made, including ensuring that it does not unfairly treat people or divulge personal information. So, the future of these music suggestion systems is about making them even better, fitting in with different cultures, and making your experience with them special and helpful in many ways.

The main contribution of this proposed model is detailed as follows:

- The suggested model records facial expressions with an integrated camera and then uses a machine learning algorithm trained on face expression datasets to analyze them.
- A CNN, a machine learning system, is used to identify emotions. It has been trained on multiple large-scale datasets.
- The recommendation of music by the ML model is based on the user's emotional state.
- To access the recommended music playlist, the model directs the user to a music streaming service.

The workflow of the framework is described in the following steps. In Sect. 2, the current module is examined and the impact of music therapy on facial emotion recognition in different patient populations is investigated. In Sect. 3, a novel approach for identifying emotions using CNN technology is presented. Next, based on the identified emotion, the model creates a playlist that may be connected to the user's favourite music provider. The experimental findings and discussion are covered in Sect. 4. Lastly, Sect. 5 provides a conclusion.

## 2 Related Works

With the use of an integrated camera, the machine learning model—which is based on SVM algorithms—captures facial expressions, runs the information through a trained machine learning algorithm, and reliably predicts emotional states, obtaining an 84.82% accuracy rate. The system has the potential to be used for music therapy, tailored song recommendations, and improving public space atmospheres because of its easy-to-use integration and lack of additional hardware requirements. Subsequent advancements may entail the use of physiological signals to augment precision in musical recommendations [2–6].

Using models like LSTM, CNN, CNN-LSTM, and LSTM-CNN, it investigates the link between Natural Language Processing (NLP) and Deep Learning, recognizing emotions like happy, sadness, wrath, and love. The model uses Long Short-Term Memory (LSTM) networks and CNN as ML approaches. The best model, a CNN, is integrated into a program that recognizes facial expressions as a type of user input. By using patterns and linkages to identify songs and playlists that match the identified emotion, the algorithm improves the experience of music recommendations [7–11].

CNNs, PCA, and SVM are three ML techniques that the system uses to train and identify human emotions. It employs techniques such as the CoAtNet algorithm, data augmentation, and the Viola-Jones algorithm to accurately identify emotions and select music that elevates the user's mood. The CoAtNet model performs more accurately than more traditional methods like CNNs, PCA, and SVM. The implementation of the STM32H747I Board demonstrates the system's practicality and verifies its effectiveness as an emotion-responsive music recommendation system [11–15].

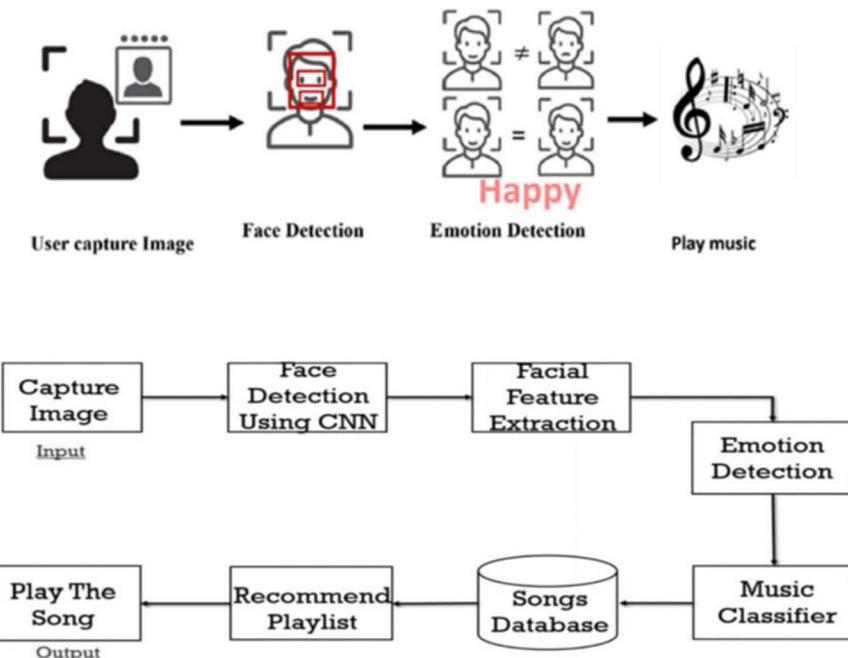
An ML model whose primary goal is to discern between five distinct emotions: shocked, furious, sad, pleased, and neutral. CNN and other ML technologies are used by the model. The algorithm recommends music depending on the specific mood recognized by analyzing facial expression patterns using CNNs [16–18]. Emo-tune uses CNNs, powerful machine learning algorithms designed for image processing and recognition. CNNs use a variety of layers, including convolutional, pooling, and fully connected layers, to effectively identify patterns and attributes associated with certain objects or classes from large labelled datasets. Once trained, CNNs can recognize new images and extract features for several applications, such as object detection and image segmentation.

### 3 Proposed System

The system utilizes an in-built camera to capture facial expressions, which are then processed using a ML algorithm trained on facial expression datasets. This algorithm predicts the person's emotional state based on their facial expressions and recommends music accordingly. Webcam image capture utilizes a device's camera to record real-time still images or video, often integrated into laptops and external devices for convenient visual information capture. This process serves various applications, including facial recognition, emotion analysis, and other computer vision tasks. Emotion classification involves identifying and categorizing the emotional state conveyed by an image or set of features, often focusing on analysing facial expressions captured through a webcam. Machine is trained on facial expression datasets to recognize patterns associated with emotions like happiness, sadness, or anger.

Emotion recognition goes beyond classifying emotions, aiming to understand and interpret emotional states. It encompasses capturing emotional cues from diverse modalities, such as facial expressions, body language, and voice tone. In webcam-based emotion recognition, the emphasis is typically on analysing facial expressions to offer a nuanced understanding of the user's emotional state beyond simple categorization. Music recommendation suggests tailored songs or playlists based on an individual's preferences, context, or emotional state. After recognizing the emotional state through webcam-based emotion analysis, a music recommendation system utilizes this information to suggest music aligned with the user's current mood. Figure 1 shows the working mechanism of the ML model. ML algorithms play a key role in learning patterns between emotional states and music preferences, delivering a personalized and context-aware music recommendation experience.

Integrating Histograms of Orientation Gradients (HOG) with ML, improves the gradient computation, and exploring multi-scale approaches. Through webcam the user's image is captured and taken as input. The emotional patterns of the facial expressions are identified by the ML algorithm CNN. The detected emotion is selected from the emotion classification.



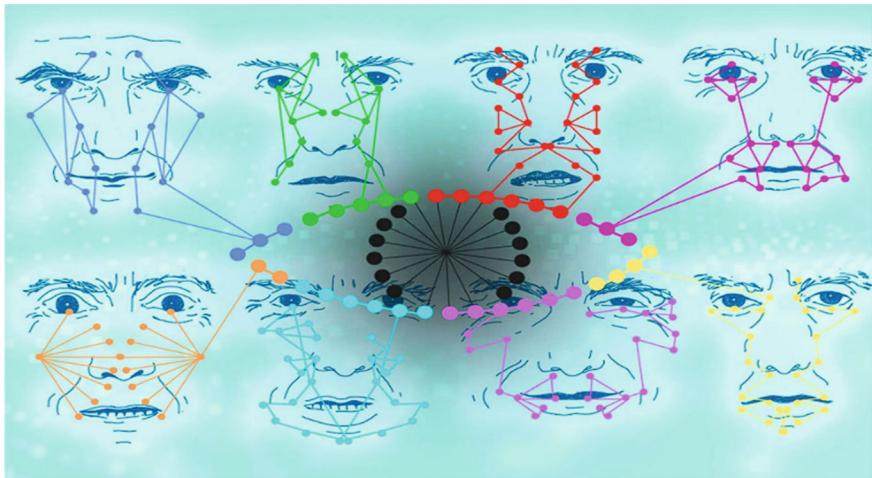
**Fig. 1** Emo-tune mechanism

The HOG is employed where through Gradient Computation the image captures by the camera it computes the horizontal and vertical gradients using convolution. Using the cell division, it divides the captured facial image into small cells. The Orientation Binning Creates a histogram for gradient directions in each cell. The Block Normalization Normalizes histograms over larger blocks to reduce illumination and contrast effects. The Feature Vector Assembly Concatenates all normalized histograms to form a feature vector. By visualization it displays the original image and the HOG feature image.

As shown in Fig. 2 the emotions are identified based on facial patterns. For each emotion, a correlated playlist is created and assigned as per user's choice of interest in any music platforms like Spotify, YouTube, Apple Music, etc. Songs are classified according to each emotion. As soon as the emotion is detected, the respective playlist from the music platform is selected by the system and it recommends the music playlist for the user.

### Formula and Equations

A model's accuracy is a summary of its predictions compared to actual values. Several accuracy metrics are derived from the confusion matrix and provide information about the model's performance. As shown in Fig. 3, some typical metrics obtained from a confusion matrix are as follows:



**Fig. 2** Emotion identification based on facial patterns

		Ground truth		
		+	-	
Predicted	+	True positive (TP)	False positive (FP)	Precision = $TP / (TP + FP)$
	-	False negative (FN)	True negative (TN)	
		Recall = $TP / (TP + FN)$	Accuracy = $(TP + TN) / (TP + FP + TN + FN)$	

**Fig. 3** Confusion matrix for accuracy prediction

- True Positive (TP): This is the situation in which the model accurately forecasts a positive outcome (such as the existence of a condition) when the real result is positive.
- True Negative (TN): This is the situation in which the model accurately forecasts a negative outcome (such as the lack of a condition) when the real result is negative.
- False Positive (FP): This occurs when a model predicts a positive result when a negative result occurs. Another name for it is a Type I mistake.
- False Negative (FN): This happens when the model predicts a negative result when a positive result happens. Another name for it is a Type II mistake.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True negative}}{\text{Total Sample}}$$

Accuracy metrics of this machine learning model is calculated as,

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\%$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TP} + \text{FP}} \times 100\%$$

$$\text{Precision} = \frac{\text{TN}}{\text{TP} + \text{FP}} \times 100\%$$

## 4 Results and Discussion

The process begins by training the machine to associate playlists with different emotions, such as happy or sad. Following this, the model undergoes testing to evaluate its accuracy in predicting emotions. The anticipated outcomes are then compared with the actual results, and any disparities lead to adjustments in the model. This iterative cycle aims to enhance the machine's ability to effectively link emotions with the appropriate playlists. This iterative approach fine-tunes the model, contributing to its proficiency in accurately catering to users' emotional preferences. This user-friendly procedure simplifies emotion-based music recommendation. Users can effortlessly navigate and engage with the system, promoting accessibility and convenience for a diverse audience. The emphasis on simplicity ensures an intuitive and enjoyable experience, allowing users to select music based on emotions without unnecessary complexities. Covering a broad spectrum of tasks, OpenCV plays a pivotal role in this paper by facilitating fundamental operations like image manipulation, filtering, and transformation, as well as supporting more sophisticated applications such as object recognition, face detection, and gesture tracking.

The basic human emotions are primarily considered and classified. Six different kinds of emotions are considered and trained on extensive datasets. The machine learning algorithm CNN works on the image recognition and facial emotion patterns of the user. Figure 4 shows how the user's emotional patterns and orientations are identified to recognise the facial expression.

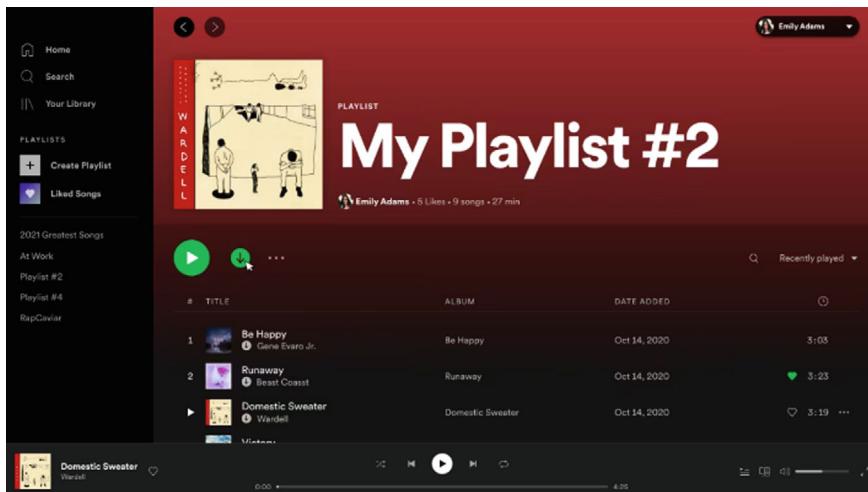
The system uses the taken image to assess emotions, and then it effortlessly directs the user to well-known music streaming services like YouTube, Spotify, or Apple Music, where it makes recommendations for songs based on the emotional state it has discovered. The user's preferences are used to categorize songs in advance, making it



**Fig. 4** Different kinds of emotion identification

possible to create custom playlists. The playlists made for every emotion are shown in Fig. 5, guaranteeing a customized listening experience.

The user is sent to a specific music playlist that is catered to the detected emotion by the system after emotion detection. By providing music choices that correspond with the emotional state of the user, this smooth integration of emotion detection and music streaming platforms enhances user engagement and happiness.



**Fig. 5** Classified playlists for each emotions

## Outcomes

**Accuracy Values:** After testing nearly 100 samples, the accuracy observed is tabulated for every evaluated emotion. An average accuracy of 86% is obtained. Table 1 shows the accuracy scores outline the performance of a machine learning model in classifying different emotions as.

### Accuracy Prediction

Notably, the model exhibits high accuracy for identifying instances associated with happiness (93%) and the rock genre (90%), showcasing its proficiency in recognizing and categorizing content related to these emotions. Additionally, the model demonstrates good accuracy for sadness (85%) and surprise (83%). However, the accuracy is relatively lower for anger (79%). From Fig. 6, 78–82% of accuracy is obtained when training and testing results are compared with each other. Overall, the results indicate the model's effectiveness in capturing various emotional states, with

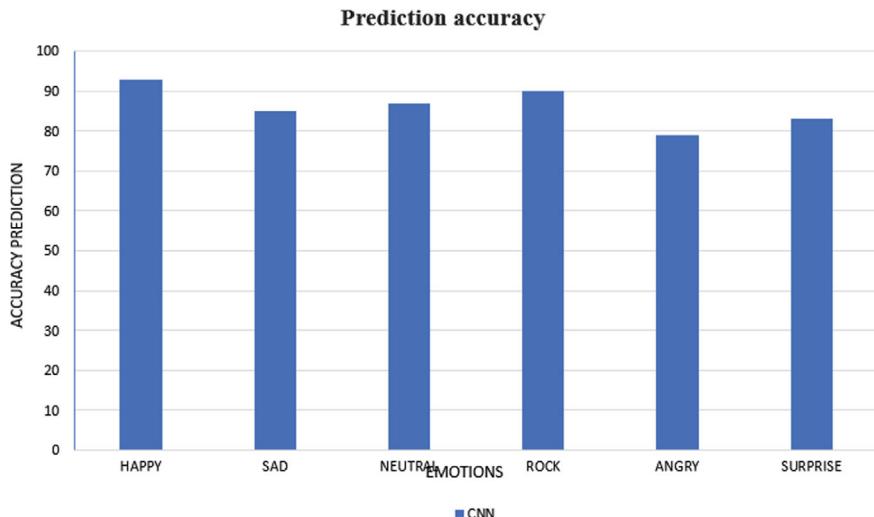
**Table 1** Accuracy scores

Emotions	Accuracy (%)
Happy	93
Sad	85
Neutral	87
Rock	90
Angry	79
Surprise	83

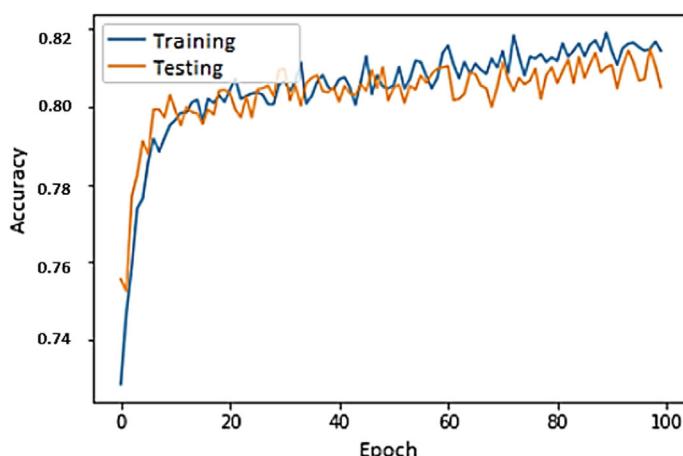
particular strength in happy and rock classifications, while also suggesting potential areas for improvement, especially in accurately identifying instances of anger.

The accuracy scores outline the performance of a machine learning model in classifying different emotions as shown in Fig. 7. The accuracy metrics and its respective scores are tabulated in Table 2.

As shown in Fig. 8 the prediction accuracy is compared between CNN and SVM algorithm. In the end, music recommendation system is designed that's easy for users to enjoy. The system is designed to suggest music based on how people are feeling.



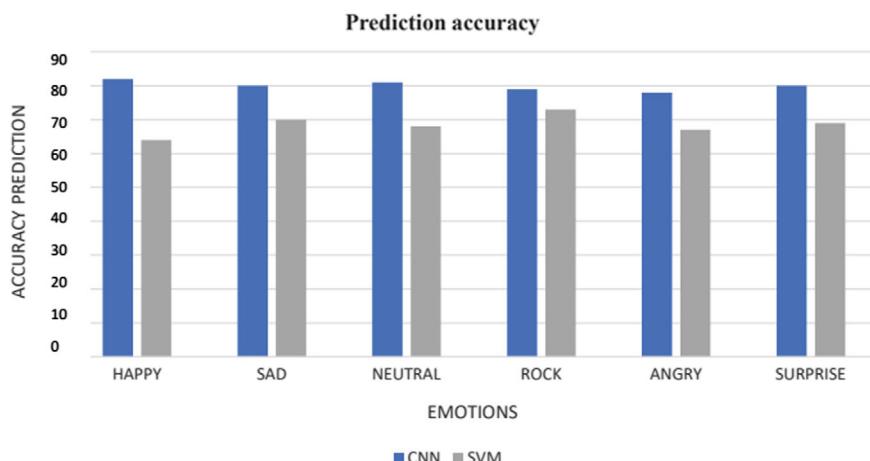
**Fig. 6** Accuracy prediction



**Fig. 7** Accuracy metrics

**Table 2** Accuracy metrics

Accuracy metrics	Scores
Accuracy	86
Recall (sensitivity)	88
Specificity	84
Precision	84.6

**Fig. 8** Comparison of accuracy prediction

It's good at figuring out what emotions people are expressing and provides accurate recommendations. Plus, it does this quickly and without any delays, so users get the music they want right away. The goal is to make sure the system is user-friendly, meaning it's easy for people to use, and it excels at understanding and responding to human emotions, giving everyone the best music experience possible.

## 5 Conclusion

In summary, 'Emo-tune: Harnessing Emotion-Based Music for Patient Well-being' is a significant advance in the use of innovative technology and helpful treatments. Smart technology is utilized to detect emotions on people's faces, and then recommends music that is just right for each individual. The special part is that it can change and choose the music therapy based on how each person is feeling, which seems like a good way to make treatments work even better. Providing healthcare that caters to individual needs while ensuring privacy and security is a beacon of hope. It plays a significant role in the ongoing transformation of healthcare practices, providing a

more personalized and emotionally resonant way to address patient wellness. ‘Emotune: Harnessing Emotion-Based Music for Patient Wellness’ has promising potential and opens up many possibilities for improvements and wider applications. Systems that suggest music based on your emotions are looking very promising in the future. These systems are improving their ability to comprehend feelings, even when there are subtle differences and how emotions are expressed in different cultures. The idea of using wearable technology like smartwatches is exciting because it allows the system to quickly adjust the music it suggests based on your current emotions. Adding additional information, such as your heart rate or brainwaves, could enhance the accuracy of the music suggestions. They are considering integrating it with other smart helpers on your devices, so it’s not just about music—it could also provide other types of support. They comprehend that there are significant considerations to be made, including ensuring that it does not unfairly treat people or divulge personal information. So, the future of these music suggestion systems is about making them even better, fitting in with different cultures, and making your experience with them special and helpful in many ways.

## References

1. Elefant C, Lotan M, Baker FA, Skeie GO (2012) Effects of music therapy on facial expression of individuals with Parkinson’s disease: a pilot study. *Music Sci* 16(3):392–400. <https://doi.org/10.1177/1029864912458917>
2. Shaila SG, Rakshita R, Shangloo A (2022) Music therapy for mood transformation based on deep learning framework. In: Computer vision and robotics, Proceedings of the CVR, Singapore, pp 35–47. [https://doi.org/10.1007/978-981-16-8225-4\\_4](https://doi.org/10.1007/978-981-16-8225-4_4)
3. Sudheesh KV, Rohit Vaidya S, Rajeeva S, Roopesh Ramalingam, Samarjith MN, Kiran (2023) A comprehensive overview on musical therapy using facial expression recognition. In: 2023 international conference on distributed computing and electrical circuits and electronics (ICDCECE), Ballar, India, pp 1–5. <https://doi.org/10.1109/ICDCECE57866.2023.10151044>
4. Rahman JS, Gedeon T, Caldwell S, Jones R, Jin Z (2021) Towards effective music therapy for mental health care using machine learning tools: human affective reasoning and music genres. *J Artif Intell Soft Comput Res* 11(1):5–20. <https://doi.org/10.2478/jaiscr-2021-0001>
5. Jaiswal S, Nandi GC (2020) Robust real-time emotion detection system using CNN architecture. *Neural Comput Appl* 32:11253–11262. <https://doi.org/10.1007/s00521-019-04564-4>
6. Burns JL, Labb   E, Arke B, Capeless K, Cooksey B, Steadman A, Gonzales C (2002) The effects of different types of music on perceived and physiological measures of stress. *J Music Ther* XXXIX(2):101–116
7. Valenzuela A, Hornero G, Royo D, Aguilar A, Casas O (2016) Emotional states through physiological signals and its application in music therapy for disabled people. *IEEE Access* 4. <https://doi.org/10.1109/ACCESS.2020.3008269>
8. Chen J et al (2020) An automatic method to develop music with music segment and long short term memory for tinnitus music therapy. *IEEE Access* 8:141860–141871. <https://doi.org/10.1109/ACCESS.2020.3013339>
9. Li D (2023) Music therapy in mental health and emotional diversion of primary and secondary school students. *Occup Ther Int* 2023:1 page. Article ID 9893830. <https://doi.org/10.1155/2023/9893830>

10. Vayadande K, Narkhede P, Nikam S, Punde N, Hukare S, Thakur R (2023) Facial emotion based song recommendation system. In: 2023 international conference on computational intelligence and sustainable engineering solutions (CISES), Greater Noida, India, pp 240–248. <https://doi.org/10.1109/CISES58720.2023.10183606>
11. Thomas M, Shreenidhi HS (2022) ANN based facial emotion detection and music selection. In: 2022 international interdisciplinary humanitarian conference for sustainability (IIHC), Bengaluru, India, pp 927–931. <https://doi.org/10.1109/IIHC55949.2022.10060593>
12. Joshi S, Jain T, Nair N (2021) Emotion based music recommendation system using LSTM-CNN architecture. IEEE Trans Neural Netw Learn Syst 32(4):1418–1432. <https://doi.org/10.1109/TNNLS.2020.2985588>
13. Gilda S, Zafar H, Soni C, Waghurdekar K (2017) Smart music player integrating facial emotion recognition and music mood recommendation. In: 2017 international conference on wireless communications, signal processing and networking (WiSPNET), Chennai, India, pp 154–158. <https://doi.org/10.1109/WiSPNET.2017.8299738>
14. Bhardwaj Y, Upadhyay A, Chauhan H, Roy NR (2022) A contemplation on music recommendation systems based on emotion detection. In: 2022 12th international conference on cloud computing, data science & engineering (confluence), Noida, India, pp 356–361. <https://doi.org/10.1109/Confluence52989.2022.9734209>
15. Ghosh O, Sonkusare R, Kulkarni S, Laddha S (2022) Music recommendation system based on emotion detection using image processing and deep networks. In: 2022 2nd international conference on intelligent technologies (CONIT), Hubli, India, pp 1–5. <https://doi.org/10.1109/CONIT55038.2022.9847888>
16. Bhatia M, Bhatia S, Hooda M, Namasudra S, Taniar D (2022) Analyzing and classifying MRI images using robust mathematical modelling. Multimed Tools Appl. <https://doi.org/10.1007/s11042-022-13505-8>
17. Agrawal D, Minocha S, Namasudra S, Kumar S (May 2021) Ensemble algorithm using transfer learning for sheep breed classification. In: IEEE 15th international symposium on applied computational intelligence and informatics, Timișoara, Romania. <https://doi.org/10.1109/SAC51354.2021.9465609>
18. Gupta A, Namasudra S (2022) A novel technique for accelerating live migration in cloud computing. Autom Softw Eng 29(34). <https://doi.org/10.1007/s10515-022-00332-2>

# Making Data Secure in Detecting ADHD with Supervised Learning



Deepak Kumar Khandelwal and Mahesh Chandra Govil

**Abstract** Attention Deficit Hyperactivity Disorder (ADHD) arises from a blend of genetic, environmental, and neurological influences. Early diagnosis and tailored treatments can greatly enhance outcomes. However, advancements in computerized detection raise privacy concerns due to unsecured data transmission. While Machine Learning (ML) enhances productivity, it encounters practical challenges. To mitigate these issues, we suggest using Privacy-Preserving Machine Learning (PPML) algorithms to secure ADHD mental health data. This involves preprocessing raw data with methods like missing value imputation and min-max normalization. The Synthetic Minority Over-sampling Technique (SMOTE) tackles class imbalance, notably ADHD patients. Following this, a differential privacy algorithm is implemented along with ML classification algorithms such as K-Nearest Neighbor (KNN), Gradient Boosting (GB), Support Vector Machine (SVM), and Random Forest (RF). Gradient Boosting (GB) stands out for its performance, with a minor accuracy compromise acceptable for ensuring data privacy. This method aims to advance ADHD understanding and management while protecting sensitive information.

**Keywords** SMOTE · Missing value replacement · Min-Max normalization · Differential privacy

## 1 Introduction

Health includes both physical and mental well-being, with mental wellness indicating a state of positive mental health that enhances social, occupational, and familial functions. ADHD, a neuropsychiatric disorder, affects approximately 5% of adults globally and 3–9% of school-aged children, leading to academic challenges, behavioral issues, relationship problems, emotional instability, and oppositional or disruptive behaviors in young individuals Loh et al. [4]. Machine learning has significantly

---

D. K. Khandelwal (✉) · M. C. Govil  
National Institute of Technology, Sikkim, Ravangla, India  
e-mail: [phcs190009@nitsikkim.ac.in](mailto:phcs190009@nitsikkim.ac.in)

contributed to the classification of ADHD by analyzing datasets. However, insecure data transmission networks pose risks such as unauthorized access and potential misuse of sensitive information, emphasizing the crucial need to protect patient privacy, especially for conditions like ADHD.

Securing medical private data through Machine Learning (ML) methods is a complex undertaking that requires robust frameworks and stringent protections. ML algorithms are increasingly used to analyze large medical datasets, such as patient records and diagnostic scans, but concerns about the confidentiality and privacy of this data remain significant. To address these issues, ML-driven approaches like differential privacy and federated learning have emerged as promising solutions. Differential privacy ensures that the outcomes of data analysis are not influenced by the presence or absence of any individual's data, thereby reducing privacy risks. Federated learning allows model training across distributed datasets without exposing raw data, preserving privacy at the data source. By enabling computations on data while maintaining the confidentiality of sensitive information, these techniques help safeguard patient privacy. Incorporating ML-driven privacy-preserving techniques allows healthcare organizations to advance medical research and patient care while maintaining the highest standards of data privacy and security.

Major contributions of the developed model are given below:

- Data pre-processing steps missing value imputation and min-max normalization are applied to ADHD data collected from the NSCH website.
- The incorporation of SMOTE addresses the issue of class imbalance by generating synthetic samples for the minority class (ADHD patients).
- Apply differential privacy to build the model to provide relevant information about ADHD dataset without releasing any personal information

The overall layout of this study is as follows: Sect. 2 provides a review of previous studies, detailing their findings and identifying any shortcomings. Section 3 describes the proposed methodology, including dataset descriptions, preprocessing methods, missing value replacement, min-max normalization, and SMOTE. Results and discussion are covered in Sect. 5, and the conclusion is presented in Sect. 6.

## 2 Literature Review

In recent times, several studies have been conducted to investigate the recognition of children's behavior. This section will delve into a review of select articles from previous research in this area.

Sharma et al. [6] developed a machine learning-based smart healthcare system aimed at automating ADHD detection. Their study extensively explores various methods for diagnosing ADHD using the ADHD200 dataset, leveraging well-established ML techniques such as neural networks and SVM models. Through their analysis, they found that phenotypic data analysis outperformed supervised learning

techniques including logistic regression, KNN, AdaBoost, among others. Furthermore, they tailored neural networks specifically using the functional connectivity of MRI data for a cohort of 40 patients with no prior neuroscience background. The integration of an ensemble classifier led to noticeable improvements in accuracy during both testing and training phases. However, despite these advancements, the efficacy of diagnosis techniques remains somewhat limited with respect to security aspect.

Zhang-James et al. [10] applied EEG time–frequency decomposition to train a convolutional neural network for classification purposes. Through this approach, an accuracy of 88% was attained, eliminating the necessity for manually selecting EEG spectral or channel features. Nonetheless, despite these positive outcomes, the study neglects to address the essential security needs of the data, leaving its privacy vulnerable.

In their work, Md. Maniruzzaman and collaborators devised a methodology aimed at identifying risk factors associated with ADHD in children and implementing a machine learning-based approach to differentiate between healthy individuals and those diagnosed with ADHD. Through a combination of oversampling and undersampling techniques, they achieved a balanced class label in their dataset. They employed eight ML-based classifiers for ADHD prediction with RF with the highest accuracy of 85.5%. Nevertheless, further enhancements are needed in the models to accurately identify risk factors linked to children diagnosed with ADHD along with security challenges.

Tor et al. [8] presented a specialized automated system tailored to assist clinicians in reaching diagnostic conclusions more effectively. Through tenfold validation using EEG data from 123 children, their study signifies the initial endeavor to refine an automated system for classifying ADHD, CD, and ADHD+CD categories based on EEG data. Notably, KNN classifier achieved an impressive accuracy of 97.88%. However, lack of addressing the security need of data as privacy of data is at stake.

Guven et al. [1] conducted research to explore a novel method for identifying ADHD patients from controls using concurrent electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS). The study included 23 children with pre-medication mixed-type ADHD and 21 healthy children. Machine learning techniques, including SVM, NB, and MLP were applied to both EEG data alone and combined fNIRS and EEG signals to determine the most accurate categorization. Naive Bayes yielded the most accurate classifications using EEG and EEG–fNIRS systems, with accuracy rates of 79.54 and 93.18%, respectively. However findings of the safeguarding the privacy of data is not considered.

Majid Moghaddari et al. [5] introduced a convolutional neural network (CNN) image processing model within the domain of deep learning. They preprocess EEG data by filtering out noise and objects, subsequently dividing it into smaller samples. Each sample undergoes the extraction of various frequency bands, resulting in the creation of RGB color images with three channels. A 13-layer CNN is then trained on these images to extract features and conduct classification tasks. However, a significant challenge was with respect to data privacy for CNN architectures.

Upon thorough analysis, it has been observed that there is a noticeable gap in research concerning the assurance of data privacy to prevent the misuse of ADHD data. Existing studies predominantly concentrate on improving accuracy and reducing error rates. However, challenges such as privacy concerns, particularly regarding ADHD data breaches, have not received adequate attention, consequently jeopardizing data integrity. To address this issue, we propose a methodology with a data quality enhancement measure incorporating missing value replacement and min-max normalization techniques integrated with differential privacy and employing five machine learning-based classifiers for analysis and prediction, aiming to identify the most accurate classifier for the task at hand.

### 3 Proposed Methodology

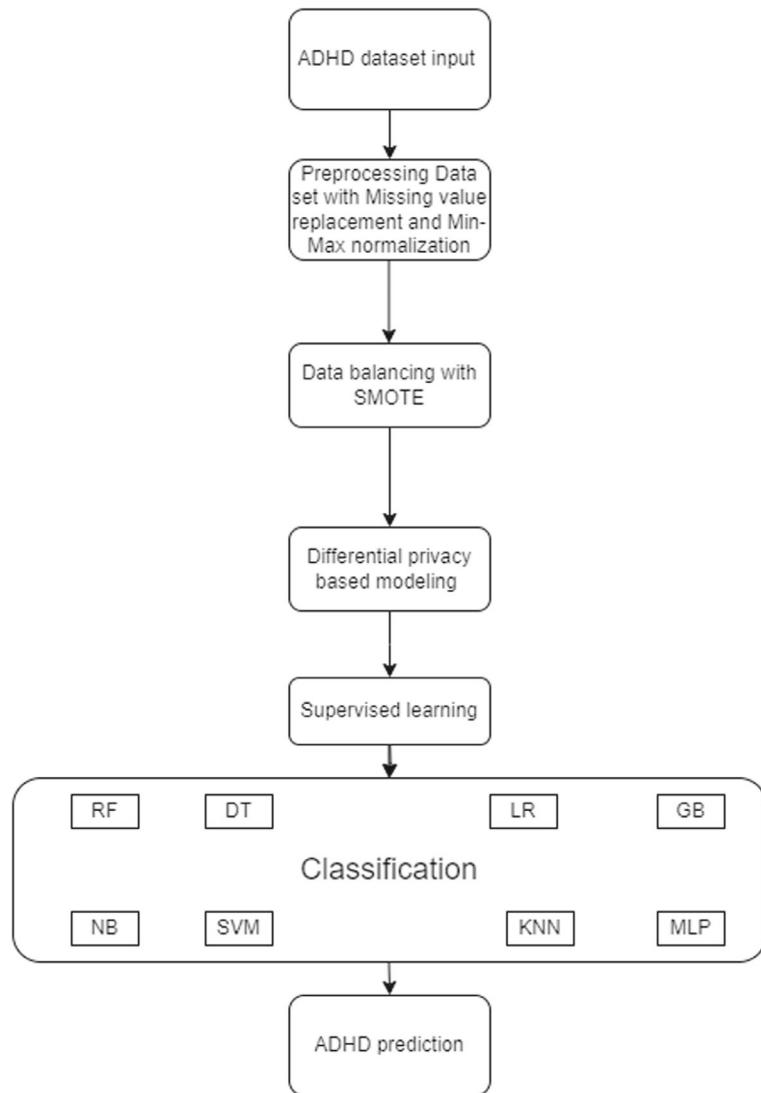
The suggested model is crafted around the supervised learning scenario for early ADHD diagnosis. Figure 1 outlines the workflow of implementing differential privacy for identifying ADHD in children, ensuring data security throughout the process.

#### 3.1 Dataset

The NSCH survey is conducted through online platforms or by mail, without the involvement of interviewers by the Health Services Dept., USA. Its primary aim is to evaluate the physical and emotional well-being of children aged 0–17 years. The NSCH pays considerable attention to various factors influencing children’s welfare, such as access to high-quality healthcare, family dynamics, parental health, neighborhood characteristics, and experiences both in school and after-school settings. Additionally, the survey seeks to assess the prevalence and impact of special healthcare needs among children. Topics addressed in the survey encompass functional limitations, transition services, shared decision-making, and satisfaction with care. Data is collected from parents or caregivers who are knowledgeable about the child’s health. The dataset comprises 59,963 entries, containing 834 features directly relevant to ADHD. To develop and evaluate a classifier intended to distinguish between children diagnosed with ADHD and those without, a carefully curated subset of behavioral features is extracted from this dataset. For training (80%) and testing (20%) the classifier to predict ADHD or healthy children, specific behavioral features are chosen from this dataset.

- Pre-processing

As depicted in Fig. 1, the ADHD data sourced from the NSCH website undergoes initial preprocessing stages. These stages involve handling missing values and applying min-max normalization to enhance the dataset’s quality. Additionally, to address data



**Fig. 1** The proposed model process flow

imbalance concerns, a further preprocessing step employing SMOTE is undertaken. Following preprocessing, the data is inputted into a joint algorithmic framework integrating Differential Privacy with a classification algorithm for detection purposes. This combined approach allows for robust and privacy-aware classification of ADHD-related behaviors within the dataset.

### **3.2 Missing Value Replacement**

Since the NSCH survey was completed via mail or online submissions without interviewer involvement, there's a likelihood of more items being left unanswered or skipped, potentially impacting population counts. Managing missing values presents a common challenge in data analysis, requiring tailored solutions based on their significance and prevalence in the dataset. When a column contains only a few missing values, omitting the affected data points might be reasonable. Conversely, if a column contains a substantial number of missing values, removing the entire column could be advisable. However, when a column with numerous missing values is crucial to the model, efforts should focus on salvaging as much information as possible. Various techniques, such as mean, median, or mode imputation, offer ways to replace missing values with statistical metrics derived from the available data. For example, in the case of the percentage of children aged 10–17 who were overweight/obese, a larger percentage of children (9.0%) did not have valid responses to items needed to ascertain their weight status compared to the previous example. Population estimates for overweight/obese children differ by about 0.9 million when accounting for missing cases (9.4 vs. 10.3 M). Mean imputation substitutes missing values with the average of observed values, median imputation employs the middle value, and mode imputation adopts the most frequent value.

$$\bar{y} = \sum y_i / n \quad (1)$$

In the given context,  $\bar{y}$  symbolizes the sample mean, whereas  $\sum y_i$  stands for the summation of each sample value (for instance, overweight). The index  $i$  represents the variable being analyzed, and  $n$  signifies the total number of elements in the sample.

### **3.3 Min-Max Normalization**

Min-Max normalization serves as a commonly used method to standardize numerical features within a predefined range, often spanning from 0 to 1. It functions by recalibrating each feature's values in relation to its minimum and maximum values, thereby placing them on a normalized scale. This technique guarantees that all features are uniformly mapped within the specified range, diminishing the impact of features with greater scales. Min-Max normalization is advantageous when managing varied feature distributions, facilitating better comparison and comprehension across different features.

$$y' = (y - \min) / (\max - \min) \quad (2)$$

Here  $y'$  is the normalized value,  $y$  depicts the raw value, max is the largest value, and the smallest value is min.

### 3.4 SMOTE

Managing data imbalances in classification tasks is crucial, especially when one class has significantly fewer samples than others. This situation can lead to biased models that struggle to accurately predict minority classes. To address this issue, SMOTE has emerged as an effective strategy. SMOTE works by generating synthetic samples for the minority class, thereby rebalancing the class distribution and improving the model's ability to learn from minority samples. During the SMOTE process, the algorithm identifies the k-nearest neighbors within the feature space for each minority class instance. It then creates new instances along the line segments connecting these neighbors. The number of synthetic instances produced depends on the specified oversampling ratio, which determines the extent of augmentation. Typically, SMOTE integration occurs during the preprocessing phase before training the machine learning model. The primary goal of using SMOTE is to mitigate class imbalance, resulting in a dataset that better represents the underlying distribution for model training. This approach enhances the model's ability to generalize across all classes, ultimately improving its predictive performance and reducing bias toward majority classes.

## 4 Differential Privacy

Differential privacy represents a critical advancement in the realm of machine learning and data privacy, particularly concerning large datasets where individual privacy is paramount. Within the domain of ADHD data management, the application of differential privacy emerges as a pivotal strategy. It operates as a meticulously crafted mathematical framework, carefully designed to uphold the confidentiality of individual data points within datasets while still facilitating robust statistical analyses. The genesis of differential privacy originates from the imperative to address the vulnerabilities inherent in linkage attacks. By ensuring that the output of a differential privacy algorithm remains largely consistent, regardless of whether an individual's data is part of the dataset, this framework effectively neutralizes attempts to glean sensitive information from the data. This standardized approach to privacy protection establishes a clear benchmark, making it significantly challenging for adversaries to infer or extract personal information about any particular individual. In essence, differential privacy serves as a cornerstone in the ongoing efforts to balance data utility with individual privacy rights, particularly in sensitive domains like ADHD medical research.

In a more formal context, we define a model  $M$  as  $\epsilon$ -differentially private provided that, for every pair of datasets  $x$  and  $y$  that vary by the data entry of a single individual, and for all events  $S$ , the subsequent condition is satisfied:

$$Pr[M(x)\epsilon S] \leq e^{\epsilon} Pr[M(y)\epsilon S] \quad (3)$$

Privacy parameter  $\epsilon$  signifies the acceptable level of privacy loss. A smaller  $\epsilon$  value indicates more robust privacy guarantees. For a low value of  $\epsilon$ , this can be approximated as

$$Pr[M(x)\epsilon S] \leq (1 + \epsilon) Pr[M(y)\epsilon S] \quad (4)$$

The derivation of differential privacy involves several crucial components: Differential privacy encompasses two primary concepts.

**Privacy Loss:** This metric assesses the degree to which an adversary can infer information about an individual by observing the output of a computation or analysis.

**Laplace Mechanism:** An indispensable mechanism employed to achieve differential privacy is the Laplace mechanism, which injects noise into computation outputs.

#### Algorithm: Pseudocode for Identifying Children's Behavior

**Input:** D = ADHD data

**Output:** Prediction of ADHD and Non-ADHD

Pre-processing:

1 - MMN = Min Max Normalization (D) // Enhance dataset quality

2 - SMOTE(D) // Balance data using SMOTE

Classification:

3 - DP = Differential Privacy Based Classification (D) Eq. (3)

4 If (DP == 0)

5       Print "ADHD"

6 Else

7       Print "Non-ADHD"

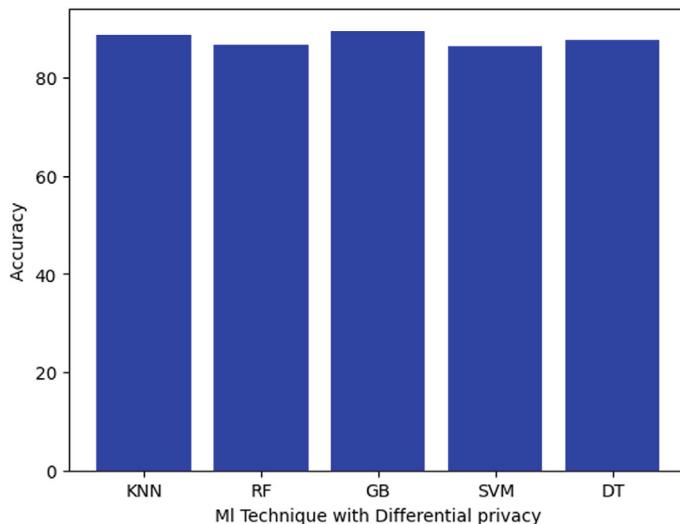
8 End

9 STOP

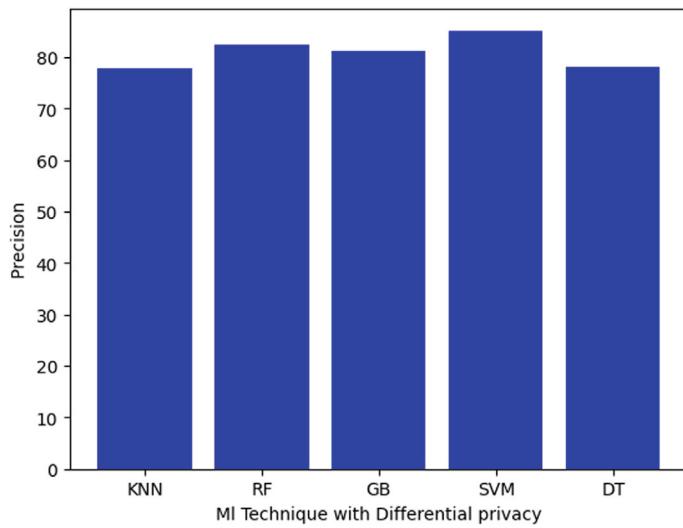
## 5 Result and Discussion

The identification of ADHD relies on supervised learning techniques. Initially, thorough behavioral data is collected from the NSCH database. Subsequently, preprocessing steps including the replacement of missing values and the application of min-max normalization techniques are executed using Python. Following this, SMOTE is employed to rectify any imbalances in the dataset. Then, the differential privacy algorithm is applied, and the applicability is evaluated through the implementation of five different machine learning algorithms KNN [9], RF [3], SVM [7], DT [2], and GB aimed at classifying features using a diverse set of learning algorithms. The assessed outcomes from the classifiers aid in determining the optimal configuration for the model's intended purpose. This model, developed through these learning scenarios, displays potential for secure ADHD identification.

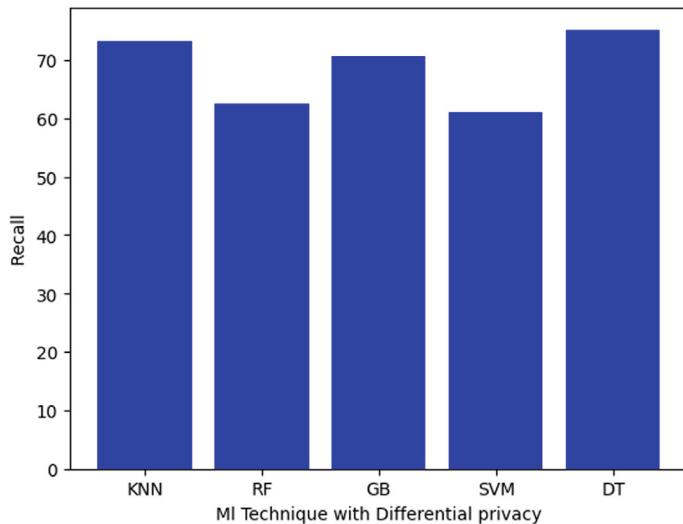
With supervised learning case in Figs. 2, 3, 4, 5, 6, 7, and 8 and Table 1 display our analysis of performance metrics across all five classifiers, offering insights into the efficacy of the designed model. We examined metrics including accuracy, error, specificity, precision, recall, and others as mentioned. In the realm of supervised learning for ADHD data supported by differential privacy, GB with differential privacy emerges as significant compared to other methodologies with relatively minimal pre-processing requirements. The top three performing algorithms in conjunction with differential privacy are GB, DT, and KNN. GB with differential privacy stands out in association with ADHD, achieving high predictive accuracy that surpasses alternative models. GB stands apart from other algorithms by adopting a principle of learning from previous model errors. It employs a stagewise addition method,



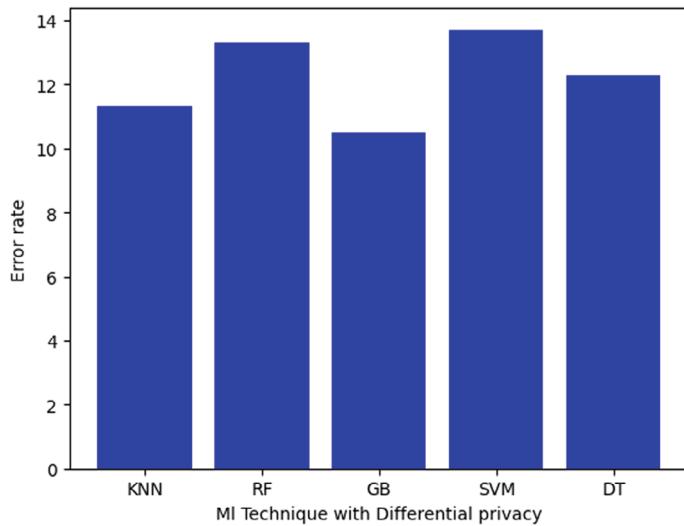
**Fig. 2** Accuracy



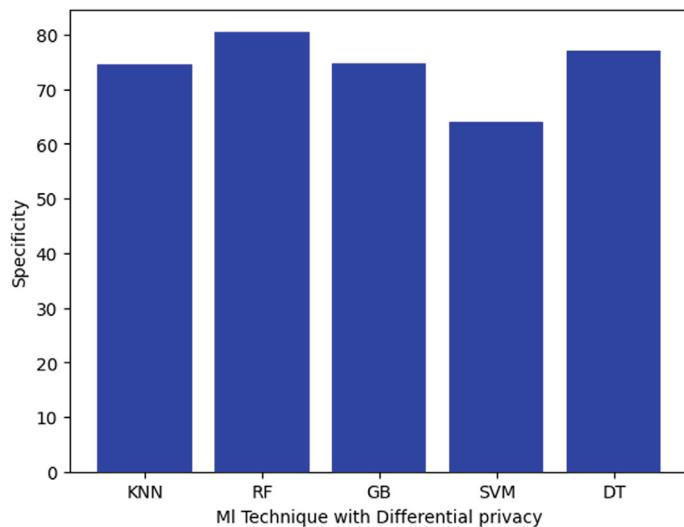
**Fig. 3** Precision



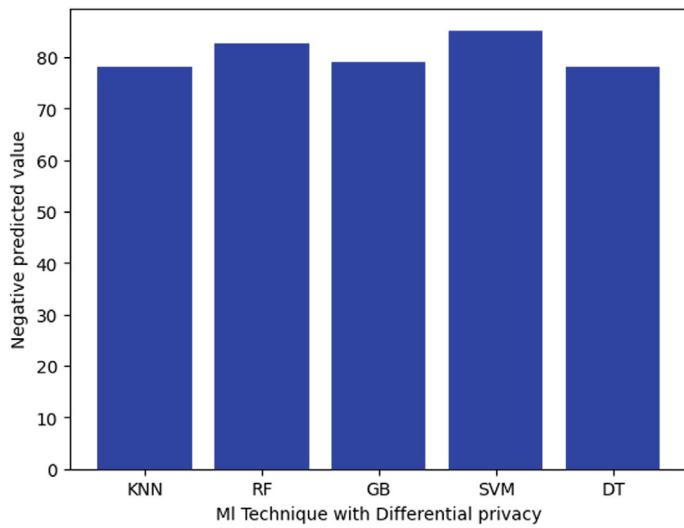
**Fig. 4** Recall



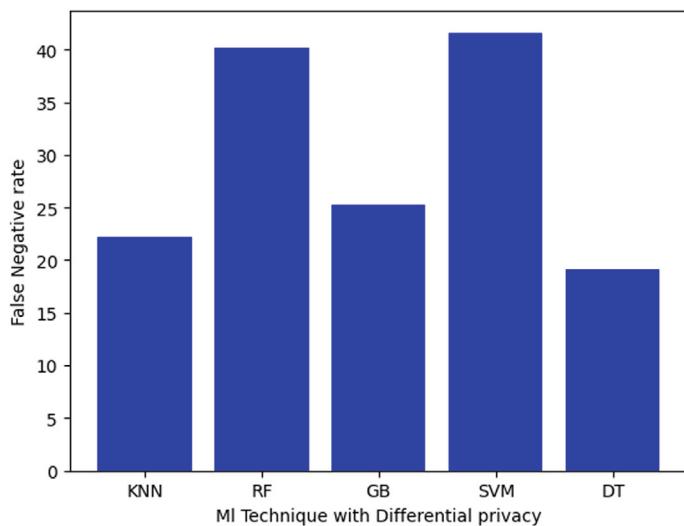
**Fig. 5** Error rate



**Fig. 6** Specificity



**Fig. 7** Negative predicted Value



**Fig. 8** False negative rate

**Table 1** Supervised machine learning with differential privacy results

	KNN	RF	GB	SVM	DT
Acc.	88.7	86.7	89.5	86.3	87.7
Pre.	77.9	82.3	81.2	85.1	78.0
Rec.	73.3	62.5	70.7	61.1	75.2
Err.	11.3	13.3	10.5	13.7	12.3
Spe.	74.4	80.5	74.7	64.0	77.0
NPV	78.0	82.5	79.1	85.1	78.0
FNR	22.2	40.2	25.3	41.6	19.1
FPR	22.1	40.3	25.2	43.9	20.7
FOR	16.8	12.3	15.7	4.8	17.9

wherein multiple weak learning algorithms are trained, and a robust learner algorithm is formed by aggregating these weak learners. Subsequently, residuals from the output of the first weak learner algorithm are calculated and utilized as the target column for subsequent weak learners. This iterative process continues until zero residuals are attained, resulting in an optimal outcome compared to other algorithms in research. GB doesn't demand comparing each new instance with an existing reference instance before determining a class, thereby enhancing accuracy with secure data, particularly in managing mixed-type ADHD data. Therefore, GB with the incorporation of differential privacy achieves a noteworthy accuracy rate of 89.5%, outperforming other competing algorithms, and underscores its aptness for the classification and security of mixed ADHD data.

## 6 Conclusion

The diagnosis of ADHD involves utilizing a structured model that incorporates the principles of differential privacy to ensure data security. This model employs supervised ML techniques. Initially, the ADHD dataset received from NSCH undergoes a data quality enhancement process which includes replacing missing values and applying min-max normalization. Subsequently, SMOTE is utilized to address data imbalance issues. The differential privacy algorithm is then applied to further secure the data. The processed datasets are evaluated using five different classifiers to predict ADHD, with the GB algorithm demonstrating the highest performance, achieving an accuracy of 89.5%. These results indicate that the GB algorithm is the most suitable for the model, ensuring secure ADHD assessment and prediction in children. Early identification and treatment of ADHD facilitated by this model can significantly benefit medical practitioners to enhance accuracy while maintaining security.

Future research could explore feature selection techniques such as Mass XGBoost. Additionally, implementing unsupervised learning with techniques like BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) for data labeling could further improve the model's performance.

## References

1. Güven A, Altunkaynak M, Dolu N, İzzetoglu M, Pektaş F, Özmen S, Demirci E, Batbat T (2020) Combining functional near-infrared spectroscopy and EEG measurements for the diagnosis of attention-deficit hyperactivity disorder. *Neural Comput Appl* 32:8367–8380
2. Hardiyanto D, Sartika DA, Rojali M et al (2019) Pedestrian crossing safety system at traffic lights based on decision tree algorithm. *Int J Adv Comput Sci Appl* 10(8)
3. Jackins V, Vimal S, Kaliappan M, Lee MY (2021) Ai-based smart prediction of clinical disease using random forest classifier and Naive Bayes. *J Supercomput* 77:5198–5219
4. Loh HW, Ooi CP, Barua PD, Palmer EE, Molinari F, Acharya UR (2022) Automated detection of ADHD: current trends and future perspective. *Comput Biol Med* 146:105525
5. Moghaddari M, Lighvan MZ, Danishvar S (2020) Diagnose ADHD disorder in children using convolutional neural network based on continuous mental task EEG. *Comput Methods Programs Biomed* 197:105738
6. Sharma A, Jain A, Sharma S, Gupta A, Jain P, Mohanty SP (2023) IPAL: a machine learning based smart healthcare framework for automatic diagnosis of attention deficit/hyperactivity disorder (ADHD). *ArXiv preprint arXiv:2302.00332*
7. Tao P, Sun Z, Sun Z (2018) An improved intrusion detection algorithm based on GA and SVM. *IEEE Access* 6:13624–13631
8. Tor HT, Ooi CP, Lim-Ashworth NS, Wei JKE, Jahmunah V, Oh SL, Acharya UR, Fung DSS (2021) Automated detection of conduct disorder and attention deficit hyperactivity disorder using decomposition and nonlinear techniques with eeg signals. *Comput Methods Programs Biomed* 200:105941
9. Xing W, Bei Y (2019) Medical health big data classification based on KNN classification algorithm. *IEEE Access* 8:28808–28819
10. Zhang-James Y, Razavi AS, Hoogman M, Franke B, Faraone SV (2023) Machine learning and MRI-based diagnostic models for ADHD: Are we there yet? *J Attent Disord* 27(4):335–353

# An Analysis of Automatic Question Generation Research Progress and Challenges



Debopam Dey and Dwijen Rudrapal

**Abstract** Automatic question generation (AQG) is a compelling and challenging area of research within Natural Language Processing (NLP). This field focuses on the automatic creation of questions from a given text, enhancing applications such as reading comprehension exercises, reducing the time teachers spend preparing questions and aiding second-language learners. The primary motivation for AQG research is the need for scalable, effective solutions to content production, evaluation and knowledge sharing. Traditional question-creation methods are labour-intensive and time-consuming, necessitating human annotators. With the rapid growth of digital data, automated systems capable of extracting relevant information and generating questions are increasingly essential. AQG aims to develop computational systems that can understand text and produce meaningful questions that test comprehension and problem-solving skills. This paper classifies various AQG approaches, analyses their results using automatic evaluation scores, reviews different datasets and their availability, and discusses current and potential evaluation techniques.

**Keywords** Automatic question generation · Machine learning · Question-answering system · Large language model · Natural language processing

## 1 Introduction

**Automatic Question Generation** (AQG) is a rapidly evolving field within Natural Language Processing (NLP) that is focused on automatically generating questions from a given text. This technology gained popularity in the early 2000s, to improve reading comprehension, saving teachers' time in creating questions and assisting second-language learners. Early AQG systems were rule-based, but modern approaches utilize deep learning and neural networks for better accuracy. Notable applications include educational tools such as Quillionz, Quizlet, Kahoot and Duolingo, which use AQG to create interactive learning activities. The future

---

D. Dey (✉) · D. Rudrapal  
Department of CSE, NIT Agartala, Agartala, India  
e-mail: [debopamdeycse19@gmail.com](mailto:debopamdeycse19@gmail.com)

of AQG looks promising with advancements in personalized learning, AI tutors and cross-domain applications, further integrating AQG into educational and professional fields.

The growing demand for scalable and effective solutions to address challenges related to content creation, assessment and knowledge sharing is the primary motivation behind AQG research. Traditional methods for generating questions often rely on manual work, where a human annotator needs to create questions manually from a given text, making it a labour-intensive, subjective and time-consuming task.

The main goal of AQG is to develop computational systems that can automatically generate questions from passages or text documents. However, the challenging aspect is to create algorithms and models capable of understanding the content of the text and generating meaningful questions that assess readers' understanding and problem-solving skills. The main contribution of this paper includes

- Analysis of the proposed promising approaches developed for AQG.
- Review of different datasets used by the researchers in this area and their availability.
- Comparative analysis of the performances of proposed AQG approaches.
- Overview of various research challenges in this domain and recent research trends.

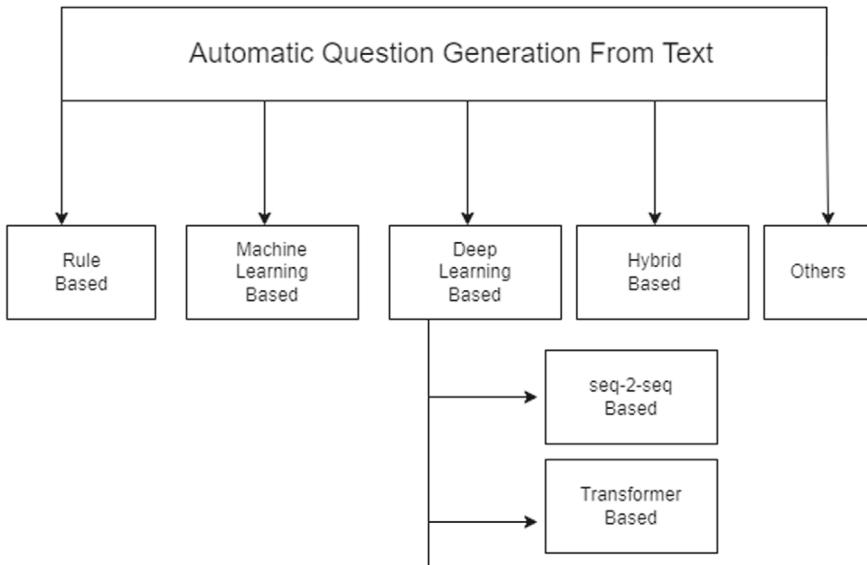
The rest of this manuscript is structured as follows. Section 2 categorizes the promising AQG approaches and elaborates key contributions followed by a comparative analysis of their performances in Sect. 3. Section 4 highlights various state-of-the-art datasets used in the AQG task. Various research challenges and recent research trends are discussed in Sects. 5 and 6 respectively. Finally, Sect. 7 concludes the survey work.

## 2 Classification of Question Generation Approaches

This literature survey categorizes AQG approaches into 5 categories, namely, Rule-based, Machine Learning-based, Deep learning-based, hybrid-based and Others. Further, the deep-learning-based approach is classified into 2 more categories, namely, seq-2-seq model-based and Transformer-based approaches. The classification structure is shown in Fig. 1.

### 2.1 Rule-Based Approaches

Rule-based models use predefined templates to extract patterns, keywords, and syntactic and semantic structures from sentences. The work in [6] proposed an approach for AQG, which used LDA for subtopic identification and ESSK for question similarity, employing syntactic tree kernels to evaluate sentence correctness. The work



**Fig. 1** Classification of different approaches used in AQG

proposed in [17] uses sentence selection, gap selection, and distractor selection, utilizing word2vec for distractor similarity, TF-IDF for text similarity, and WordNet for semantic similarity. In later works [15], content selection techniques leveraging domain knowledge, bootstrapping for semantic roles and Open Information Extraction for key semantic relations are used to generate significant questions.

## 2.2 *Machine Learning-Based Approaches*

Machine learning approaches use a large dataset to generate questions from documents, they aren't dependent on predefined rules or patterns. The work in [2] proposed an approach in 2022, presenting a framework for automatic question generation from unstructured English text, combining traditional linguistic approaches with machine learning. Lala Septem Riza et al., proposed in [26] an approach in 2023, that used simple sentence extraction from reliable news articles, followed by problem classification to determine question formats.

## 2.3 Deep Learning-Based Approaches

Deep learning provides semantic understanding, flexibility in handling variable-length input sequences, attention mechanisms for focusing on pertinent information and embedding representations for capturing semantic similarities by leveraging neural network architectures such as sequence-to-sequence (Seq2Seq) models and Transformers. The following subsection extends the promising AQG models using sequence-to-sequence (Seq2Seq) models and the Transformers model.

### 2.3.1 Sequence-To-Sequence Model

Seq-2-seq models are commonly used in neural network architecture in Deep Learning and NLP. These models consist of two parts Encoder and Decoder. Multiple numbers of RNN and LSTM can be used for both the Encoder and Decoder parts. The work by [10, 35] uses an attention-based sentence Encoder (Bidirectional LSTM) that takes the input sentence/paragraph, and then a Decoder(Bidirectional LSTM) to decode the sentence and generates the next words for question generation. A coverage mechanism to prevent word repetition problems is proposed in [5] in addition to a simple RNN encoder-decoder for question generation. The works in [19, 36] use a gated self-attention-based sentence Encoder (Bidirectional RNN) that takes the input paragraph and the answers. In the encoder part, 2 encoders namely sentence encoder and relational encoder are used in addition to word embeddings, POS embeddings and NER embeddings. In later approaches [27], RNN-based encoder-decoder architecture along with copy, attention and coverage mechanism is also utilized for the task. In a recent work in [22], authors proposed one refined neural network called Refnet with a dual attention mechanism, for Embeddings they used Word’s Globe Embeddings and in the Encoder and Decoder part they used Bidirectional LSTM. In the decoder part, they used two decoders named preliminary decoder and refinement decoder.

### 2.3.2 Transformer-Based Model

Transformer-based models in Automatic Question Generation (AQG) utilize advanced neural architectures like BERT and GPT to generate high-quality questions. These models excel in capturing context and semantics due to their self-attention mechanisms. The model proposed in [7] used an architecture based on the pre-trained BERT language model for question generation tasks. Contextualized word embeddings and pre-trained transformer model BERT for question generation tasks is used in [30] and leverages BERT’s self-attentions for language modelling of questions conditioned on paragraphs and answers phrases, implementing a copy mechanism over self-attention. The work in [34] proposed a multi-question generation model (mQG) that produces multiple, diverse, and answerable questions by focusing on context

and employing maximum question similarity loss (LMQS) and a recursive referencing process. Another work in [20] tackles factual inconsistencies and incorrect entities in question generation methods based on pre-trained language models. The proposed solution involves a data processing technique based on de-lexicalization, ensuring consistent question generation across different domains. The recent work in [24] combines an encoder-decoder-based Large Language Model (LLM) architecture with a hill-climbing algorithm to generate diverse and high-fidelity questions, balancing these aspects through a sub-modular objective function.

## 2.4 *Hybrid Approaches*

Hybrid-based models in Automatic Question Generation (AQG) combine rule-based techniques with neural network models to enhance question generation. These models leverage the strengths of both approaches: the precision of rule-based systems and the adaptability of neural networks. Approaches like [14] combines rule-based techniques and a Seq2Seq neural network model for automatic question generation, the work [31] integrates fact-based information from knowledge graphs and work in [1] utilizes ProphetNet and UniLM, fine-tuned with temporal questions generated from predefined templates. These models are trained in a sequence-to-sequence learning setup, with input comprising the sentence and candidate answer, and outputting the corresponding question.

## 2.5 *Other Approaches*

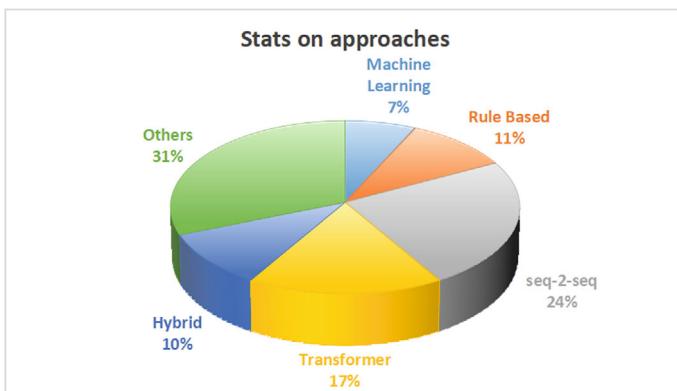
Other than the above categories, researchers also have proposed many approaches for AQG which were shown remarkable performance and outperform many state-of-the-art approaches. The work in [3] proposes a semi-autoregressive model that generates question groups in parallel. The model utilizes Transformer architecture and dual-graph interaction to enhance information extraction from passage and answer graphs. Some other approaches like the work in [9] employs a contrastive learning strategy using both ‘positive’ and ‘negative’ document sets. MSCQG consists of a document-specific generator fine-tuned from GPT-2 and an inter-document coordinator trained via reinforcement learning, the work [13] presents IGND, an Iterative Graph Network-based Decoder, to improve natural question generation (QG) from passages, the framework [8] involves constructing a context graph, selecting a reasoning chain, generating an initial question and iteratively rewriting it to increase complexity. Some of the recent approaches proposed in [4] fine-tune a pre-trained seq2seq model (BART-large) and employ methods to enhance controllability and faithfulness by steering questions towards salient keywords, Zhao et al. [37] uses a pre-trained transformer-based sequence-to-sequence model, the approach enhances children’s literacy through question-type prediction and event-centric summaries,

Naeiji [21] used Semantic Role Labeling (SRL) to transform training examples into semantic representations, enhancing the generation of diverse questions from input sentences. By incorporating SRL, the models can effectively capture sentence structures and it can generate complex questions. The very recent work in [29], Leverages item response theory (IRT), the approach generates question-answer pairs with controllable difficulty levels aligned with learners' abilities.

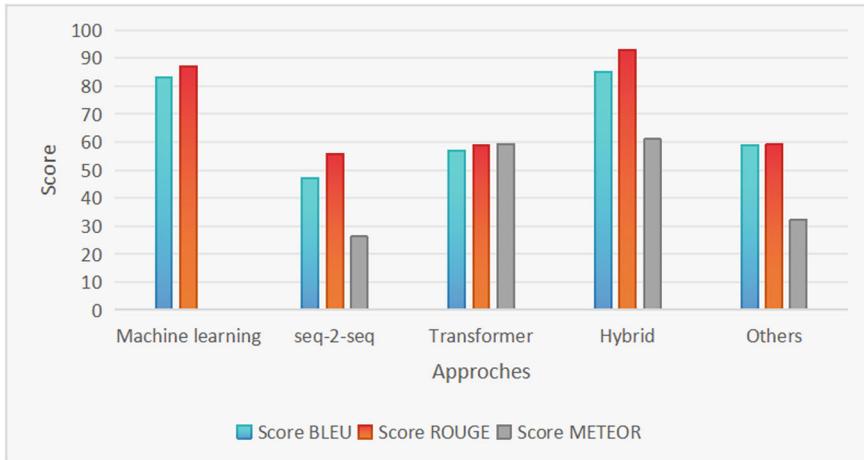
### 3 Performance Analysis of the Proposed Approaches

In this survey, we have analysed a total of 30 papers from 2015 to 2023. Total rule-based approaches were used at 11%, Machine Learning-based approaches were used at 7%, seq-2-seq approaches were used at 24%, transformer-based approaches were used at 17%, hybrid approaches were used at 10% and others approaches were used at 31%. The pie chart in Fig. 2 shows the distribution of our survey of different approaches used by the researchers for the automatic question generation task.

For this survey, we took automatic evaluation metrics BLUE, METEOR and ROUGE scores except the rule-based approach because it doesn't use these metrics for evaluating their models. In-depth analysis of reported evaluation scores for the proposed approaches as shown in Fig. 3 reveals that the machine learning approach's highest BLUE and ROUGE scores obtained are 83 and 87% respectively, for seq-2-seq-based models highest BLUE, ROUGE and METEOR scores obtained are 47.27, 55.9, 26.2%, for transformer-based models highest BLUE, ROUGE and METEOR scores obtained are 57, 58.9, 59.37%, for hybrid-based models highest BLUE, ROUGE and METEOR scores obtained are 85, 93, 61%, for others models highest BLUE, ROUGE and METEOR scores obtained are 58.9, 59.1, 32.3%. So we can see that Hybrid-based models get the highest scores for all matrices among



**Fig. 2** Distribution of various AQG approaches



**Fig. 3** Utilization of different evaluation matrices across approaches

all the approaches. All the BLUE, ROUGE and METEOR scores are quite similar for the transformer-based model. Transformer-based models get higher scores than seq-2-seq models. Other models' scores are similar to the transformer-based model but the other model's METEOR score is much less than the METEOR score of the transformer-based model.

## 4 Datasets

In this section, we provided a comprehensive overview of the datasets commonly used in **Automatic Question Generation**. We have categorized each dataset based on its domain, size, and language and reported in Fig. 4. These datasets are commonly used for training and evaluating Question generation models.

1. Stanford Question Answering (SQuAD) Dataset: SQuAD dataset [25] was introduced in 2016 primarily for reading comprehension and question answering. It consists of over 100,000 question-answer pairs taken from more than 500 Wikipedia articles. It is one of the largest and most widely used datasets for automatic question generation and answering, written in the English language.
2. Microsoft Machine Reading Comprehension (MS MARCO) Dataset: This dataset [23] was introduced in the year 2016 primarily for reading comprehension and question answering. It consists of 1.2 million real-world questions and each question corresponds to a passage. The passage and the questions in the MS MARCO dataset are written in the English language.
3. NewsQA Dataset: NewsQA dataset [28] was introduced in the year 2016 primarily for question-answering. It consists of 100000 human-generated question-

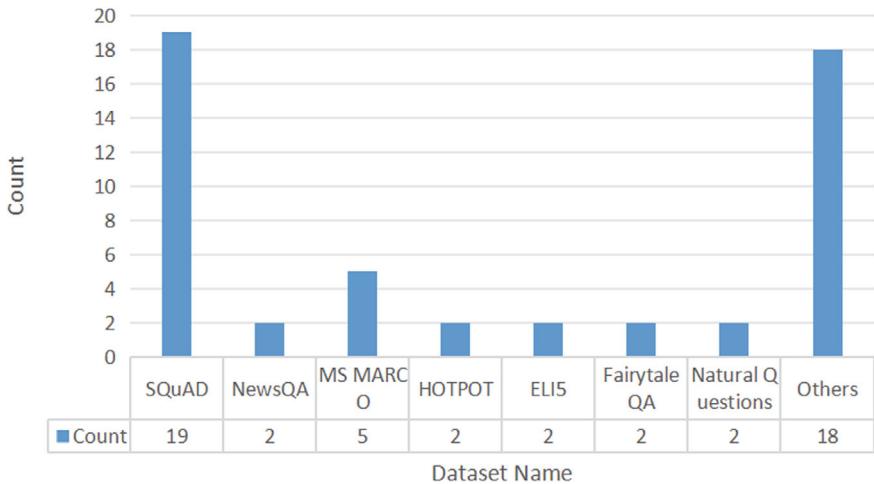
answer pairs over 10,000 news articles including politics, sports, entertainment and world events, including fact-based, opinion-based and inference-based questions, written in the English language.

4. HotpotQA Dataset: HotpotQA dataset [33] was introduced in the year 2018 primarily for question-answering. It consists of over 113 k Wikipedia-based factoid question-answer pairs, written in the English language.
5. ELI5 Dataset: ELI5 dataset [12] was introduced in the year 2019 primarily for long-form question-answering. It consists of 270 k threads from the Reddit forum, written in the English language.
6. Natural Questions Dataset: Natural Questions dataset [18] was introduced in the year 2019 primarily for question-answering. It consists of 321k real-world annotated questions collected from Google search results. The questions and the passages are written in the English language. Additionally, it has fact-based queries and opinion-based queries.
7. FairytaleQA Dataset: Natural Questions dataset [32] was introduced in the year 2022 primarily for question-generation tasks in the educational domain. It consists of 10.5 k real-world human-generated questions collected from 278 children-friendly stories, written in the English language.
8. Others Dataset: Other than the above-mentioned popular AQG dataset, some research-specific datasets also developed and utilized in various AQG research works like the dataset used in [26] have prepared a total of 1010 questions from various news domains like health, hoaxes, holidays, the environment and sports. The dataset used in [15] is prepared for 200 k company profiles. Similarly, the datasets like TriviaQA [16] and SEARCHQA [11] are also some widely used datasets for AQG tasks.

A comparative analysis of the above-discussed dataset utilization in various proposed AQG approaches is reported in Fig. 5.

Dataset Name	Domain	Number of QA pairs	Language
SQuAD	Wikipedia	100k	English
NewsQA	News articles	100k	English
MS MARCO	Bing search results	1200k	English
HOTPOT	Wikipedia	113k	English
ELI5	Reddit forum	270k	English
FairytaleQA	Children-friendly stories	10.5k	English
Natural Questions	Wikipedia and Google Search	321k	English
Others	Online articles		English, Hindi

**Fig. 4** Summary of standard datasets



**Fig. 5** Statistics on datasets used by different AQG approaches

## 5 Research Challenges

Even though automatic question creation has advanced significantly, there are still several important issues that need to be resolved. Creating thoughtful, relevant questions that accurately gauge a reader's comprehension of the assigned material is one of the main obstacles. Current algorithms frequently fail to grasp the subtleties of language and provide pertinent, grammatically sound and semantically consistent queries. Furthermore, further study and improvement are needed in the field of creating questions that focus on various cognitive levels, such as remembering, comprehending and applying. Last but not least, the absence of common assessment criteria and standards impedes the progress of the field by making it challenging to evaluate the effectiveness of various AQG systems.

- Generating questions that are contextually relevant to a given text is crucial for assessing a reader's comprehension accurately. It is a challenging task because it requires understanding the content and context of the text comprehensively.
- Maintaining coherence and correctness while generating questions automatically is not straightforward, especially when dealing with complex language structures or ambiguous text passages.
- Questions should target various cognitive levels to comprehensively assess the reader's understanding. These levels include remembering factual information, understanding concepts, applying knowledge to new situations, analysing information, evaluating arguments and creating new ideas.
- The lack of standardized evaluation metrics and benchmarks poses a significant challenge in assessing the performance of AQG systems objectively. Without reli-

able evaluation criteria, it becomes difficult to compare different systems or track improvements over time.

## 6 Current Research Trends

The field of automatic question generation is rapidly evolving, with researchers exploring various avenues to address the existing challenges. One emerging trend is the incorporation of deep learning techniques, such as sequence-to-sequence models and transformer-based architectures, which have demonstrated the ability to generate more coherent and contextual questions. Additionally, there is a growing interest in developing multimodal AQG systems that can leverage both textual and visual information to generate more comprehensive and engaging questions.

- Deep learning methods, particularly sequence-to-sequence, for instance, operate by mapping input sequences (such as text passages) to output sequences (questions) directly. Transformer-based architectures, exemplified by models like BERT and GPT (such as GPT-3), leverage attention mechanisms to capture long-range dependencies in text, enabling more nuanced question generation.
- Integrating textual and visual information in AQG systems is gaining traction. By incorporating images, diagrams or other visual aids along with textual inputs, these systems can generate questions that leverage both modalities, resulting in more comprehensive and engaging assessments of comprehension.
- Reinforcement learning (RL) and adversarial training techniques are being explored to improve the quality and diversity of generated questions. RL frameworks allow AQG models to learn from feedback received based on the relevance and quality of generated questions.
- By leveraging knowledge from related tasks or domains, transfer learning and few-shot learning may expedite model training and enhance AQG model performance in real-world applications.

## 7 Conclusion

In conclusion, automatic question generation (AQG) is at the forefront of technological innovation with significant applications in education, assessment and conversational AI. This study provides a comprehensive overview of AQG, including its motivations, existing approaches, challenges and trends. As AQG progresses, developing high-quality, contextual question-generation systems is essential for enhancing learning, assessment and interactive knowledge acquisition. These systems can personalize education, foster critical thinking and improve engagement. Future advancements depend on collaborative efforts across natural language processing, machine

learning, cognitive science and education research. AQG holds the promise to revolutionize education, making it more personalized, interactive and accessible.

## References

1. Bedi H, Patil S, Palshikar G (2021) Temporal question generation from history text. In: Proceedings of the 18th international conference on natural language processing (ICON), pp 408–413
2. Blšták M, Rozinajová V (2022) Automatic question generation based on sentence structure analysis using machine learning approach. *Nat Lang Eng* 28(4):487–517
3. Chai Z, Wan X (2020) Learning to ask more: Semi-autoregressive sequential question generation under dual-graph interaction. In: Proceedings of the 58th annual meeting of the association for computational linguistics, pp 225–237
4. Chakrabarty T, Lewis J, Muresan S (2022) Consistent: Open-ended question generation from news articles. ArXiv preprint [arXiv:2210.11536](https://arxiv.org/abs/2210.11536)
5. Chali Y, Baghaee T (2018) Automatic opinion question generation. In: Proceedings of the 11th international conference on natural language generation, pp 152–158
6. Chali Y, Hasan SA (2015) Towards topic-to-question generation. *Comput Linguist* 41(1):1–20
7. Chan YH, Fan YC (2019) A recurrent Bert-based model for question generation. In: Proceedings of the 2nd workshop on machine reading for question answering, pp 154–162
8. Cheng Y, Li S, Liu B et al (2021) Guiding the growth: difficulty-controllable question generation through step-by-step rewriting. ArXiv preprint [arXiv:2105.11698](https://arxiv.org/abs/2105.11698)
9. Cho WS, Zhang Y, Rao S et al (2019) Contrastive multi-document question generation. ArXiv preprint [arXiv:1911.03047](https://arxiv.org/abs/1911.03047)
10. Du X, Shao J, Cardie C (2017) Learning to ask: neural question generation for reading comprehension. ArXiv preprint [arXiv:1705.00106](https://arxiv.org/abs/1705.00106)
11. Dunn M, Sagun L, Higgins M, et al (2017) SearchQA: a new Q&A dataset augmented with context from a search engine. ArXiv preprint [arXiv:1704.05179](https://arxiv.org/abs/1704.05179)
12. Fan A, Jernite Y, Perez E et al (2019) Eli5: long form question answering. ArXiv preprint [arXiv:1907.09190](https://arxiv.org/abs/1907.09190)
13. Fei Z, Zhang Q, Zhou Y (2021) Iterative GNN-based decoder for question generation. In: Proceedings of the 2021 conference on empirical methods in natural language processing, pp 2573–2582
14. Flor M, Riordan B (2018) A semantic role-based approach to open-domain automatic question generation. In: Proceedings of the thirteenth workshop on innovative use of NLP for building educational applications, pp 254–263
15. Jin Y, Le P (2016) Selecting domain-specific concepts for question generation with lightly-supervised methods. In: Proceedings of the 9th international natural language generation conference, pp 133–142
16. Joshi M, Choi E, Weld DS et al (2017) TriviaQA: a large scale distantly supervised challenge dataset for reading comprehension. ArXiv preprint [arXiv:1705.03551](https://arxiv.org/abs/1705.03551)
17. Kumar G, Banchs RE, D’Haro LF (2015) RevUP: automatic gap-fill question generation from educational texts. In: Proceedings of the tenth workshop on innovative use of NLP for building educational applications, pp 154–161
18. Kwiatkowski T, Palomaki J, Redfield O et al (2019) Natural questions: a benchmark for question answering research. *Trans Assoc Comput Linguist* 7:453–466
19. Li J, Gao Y, Bing L et al (2019) Improving question generation with to the point context. ArXiv preprint [arXiv:1910.06036](https://arxiv.org/abs/1910.06036)
20. Maheshwari H, Shekhar S, Saxena A et al (2023) Open-world factually consistent question generation. *Find Assoc Comput Linguist: ACL* 2023:2390–2404
21. Naeiji A (2022) Question generation using sequence-to-sequence model with semantic role labels

22. Nema P, Mohankumar AK, Khapra MM et al (2019) Let's ask again: refine network for automatic question generation. ArXiv preprint [arXiv:1909.05355](https://arxiv.org/abs/1909.05355)
23. Nguyen T, Rosenberg M, Song X et al (2016) MS MARCO: a human-generated machine reading comprehension dataset
24. Puranik V, Majumder A, Chaoji V (2023) Protege: Prompt-based diverse question generation from web articles. Find Assoc Comput Linguist: EMNLP 2023:5449–5463
25. Rajpurkar P, Zhang J, Lopyrev K et al (2016) Squad: 100,000+ questions for machine comprehension of text. ArXiv preprint [arXiv:1606.05250](https://arxiv.org/abs/1606.05250)
26. Riza LS, Firdaus Y, Sukamto RA et al (2023) Automatic generation of short-answer questions in reading comprehension using NLP and KNN. Multimedia Tools Appl 82(27):41913–41940
27. Sasazawa Y, Takase S, Okazaki N (2019) Neural question generation using interrogative phrases. In: Proceedings of the 12th international conference on natural language generation, pp 106–111
28. Trischler A, Wang T, Yuan X et al (2016) NewSQA: a machine comprehension dataset. ArXiv preprint [arXiv:1611.09830](https://arxiv.org/abs/1611.09830)
29. Uto M, Tomikawa Y, Suzuki A (2023) Difficulty-controllable neural question generation for reading comprehension using item response theory. In: Proceedings of the 18th workshop on innovative use of NLP for building educational applications (BEA 2023), pp 119–129
30. Varanasi S, Amin S, Neumann G (2020) CopyBERT: a unified approach to question generation with self-attention. In: Proceedings of the 2nd workshop on natural language processing for conversational AI, pp 25–31
31. Wang S, Wei Z, Fan Z et al (2020) PathQG: neural question generation from facts. In: Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP), pp 9066–9075
32. Xu Y, Wang D, Yu M et al (2022) Fantastic questions and where to find them: FairytaleQA—an authentic dataset for narrative comprehension. ArXiv preprint [arXiv:2203.13947](https://arxiv.org/abs/2203.13947)
33. Yang Z, Qi P, Zhang S, et al (2018) HotpotQA: a dataset for diverse, explainable multi-hop question answering. ArXiv preprint [arXiv:1809.09600](https://arxiv.org/abs/1809.09600)
34. Yoon H, Bak J (2023) Diversity enhanced narrative question generation for storybooks. ArXiv preprint [arXiv:2310.16446](https://arxiv.org/abs/2310.16446)
35. Yuan X, Wang T, Gulcehre C et al (2017) Machine comprehension by text-to-text neural question generation. ArXiv preprint [arXiv:1705.02012](https://arxiv.org/abs/1705.02012)
36. Zhao Y, Ni X, Ding Y et al (2018) Paragraph-level neural question generation with maxout pointer and gated self-attention networks. In: Proceedings of the 2018 conference on empirical methods in natural language processing, pp 3901–3910
37. Zhao Z, Hou Y, Wang D et al (2022) Educational question generation of children storybooks via question type distribution learning and event-centric summarization. arXiv preprint [arXiv:2203.14187](https://arxiv.org/abs/2203.14187)

# A Federated Learning Approach Towards a Privacy-Preserving Technique for Brain Tumor Classification



Anurag De, Gautam Pal, Karnam Shyam, and Kalakanti Pawan Tej

**Abstract** The brain is one of the main organs in the human body. It controls almost all the actions of a human being. Any problem with the brain can lead to even fatal consequences. One of the most dangerous problems faced by the human brain is a brain tumor. If it is not treated properly at early stages, it may lead to severe conditions. However, finding a brain tumor is a very challenging task as the brain structure is very complex. Also, a lot of image data is required to train the model to detect brain tumors. However, the hospitals are not ready to share patient data to train the model. Hence, an approach to train the model by seeing the actual dataset or simply using decentralized data from multiple clients needs to be followed. So, a federated learning methodology is utilized to train the model. The trained federated learning model has achieved a training accuracy of 91% and a test accuracy of 88%. The federated learning and transfer learning approach stands out in preserving privacy, optimizing model performance, and leading to good collaboration among healthcare entities.

**Keywords** Federated learning · Transfer learning · Flower framework · Brain MR images

---

A. De (✉) · K. Shyam · K. P. Tej

School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh, India

e-mail: [anurag.de111@gmail.com](mailto:anurag.de111@gmail.com)

K. Shyam

e-mail: [karnam.shyam2004@gmail.com](mailto:karnam.shyam2004@gmail.com)

K. P. Tej

e-mail: [kpavantej09@gmail.com](mailto:kpavantej09@gmail.com)

G. Pal

Computer Science and Engineering Department, TIT Narsingarh, Agartala, Tripura, India

e-mail: [gautamtit2008@gmail.com](mailto:gautamtit2008@gmail.com)

## 1 Introduction

The human brain is one of the most essential organs of the human body and very complex to understand its internal structure. It mainly controls memory, vision, thought, emotion, breathing, and so on. It plays an important role in every process that regulates our body. So, the brain must be always protected. But brain tumors are one of the most dangerous problems, which the brain can affect badly. Brain tumor refers to the abnormal growth of brain cells in or near brain tissue. Brain tumors can generally begin in the brain and are called primary brain tumors. Later, if not treated properly at early stages, it spreads from the brain to other parts of the body. These tumors are called secondary brain tumors. A report from cancer.net [1] said that the last 5-year relative survival rate for people younger than age 15 is about 75%. For the people whose age is from 15 to 39, the 5-year relative survival rate nears 72%. The 5-year relative survival rate for people aged 40 and older is 21%. On average, the total survival rate of brain tumor patients is 85%. This survival rate can be increased even more if the brain tumor is detected in the early stages only.

For image classification, the traditional machine learning or deep learning approach can be used. Features extracted for the training data are one of the most important things while training an image classification model. But if machine learning is used, the features need to be extracted manually from the train images and then feed them to the model for training. This may take a lot of computation time to extract features from the train data.

To solve these issues, deep learning techniques and Convolution Neural Networks (CNNs) play a significant role in extracting the features from the train data. Using convolution layers, the features from the training images can be efficiently extracted. But these methods also have some disadvantages that is, even though the feature extraction was made easy we need to build a very deep neural network to train the model properly and require to select from initial weights for the model to get better outcomes.

However, the transfer learning technique can be used to train the model by using a pre-trained model [2]. The pre-trained models already have their weights set, hence training the model from scratch is not required. Because we just need to update the pre-trained model weights with the custom dataset. And it takes less time to train the model as the weights are already set. However, any approach is used, the main problem is to collect the data. In the healthcare department, it is even more difficult to collect the data. Because the healthcare departments are not ready to share their patients' data, which is confidential, to the outer world or for training purposes. The data available at each hospital is not at all sufficient to build a robust model.

So, to train a robust model to detect brain tumors, very large real-time data is the necessity. However, it is difficult to gather that huge amount of data as the healthcare departments are not ready to share their patient's personal or health details with the outer world. The primary contributions of the proposed methodology are:

- To train a model that is capable of learning patterns from decentralized data. Since patient privacy concerns prevent clinics and hospitals from sharing their data, the idea of distributed learning and federated learning (FL) is introduced.
- Training is carried out locally at each client or hospital, and only the training parameters are shared between them. Simultaneously, patient information is never shared which leads to preserving the privacy of the patients' vital health information.

Therefore, to solve this issue, the federated learning methodology described above is employed. One of the subfields of machine learning [3] called "federated learning" or "collaborative learning" focuses primarily on training a model cooperatively while maintaining the decentralization of client data.

The rest of the paper has been organized in the following order: A comprehensive review of the existing techniques in the chosen field of research is provided in Sect. 2. The introduced federated learning approach is explained in Sect. 3. Section 4 discusses the experimental results. The paper's conclusion and potential for future improvements are described in Sect. 5.

## 2 Related Work

The research and literature that is currently available on the chosen field of research is compiled in this section.

Authors in [4] applied the hybrid methodology that combines pre-trained Google Net architecture with a CNN model and changed some layers to improve training. With this approach, they overcome the challenges of manual feature extraction. As expected, the trained model outperformed transfer learning models and different machine learning/DL models on the Kaggle dataset, achieving high accuracy, Precision, Recall, and F1-Score. In [5] the authors proposed a methodology that uses a segmentation approach to identify the infected region to analyze the severity of the tumor. The researchers used the inceptionv3 model to extract deep features, and then it was given to the QVR (quantum vibrational classifier) for further classification. The trained model achieved over 90% detection scores on three benchmark datasets. The study was evaluated on MATLAB 2021 RA and a Windows 10 operating system. Authors in [6] aim to find the most effective method for calculating the brain tumor area in MRI images and to find the number of tumors in the given area. The authors used various methods, like Fuzzy C-Means, Herbaceous Method, Region Growing, and Self Organizing Maps, to analyze MRI images. They also used advanced imaging techniques to separate tumor spots and brain fluids. They found that FCM is the best method for tumor brain detection. The tumor zone was detected, and the volume of the affected region was calculated by using techniques some popular techniques like Watershed, Graph-Cut, and Active Counter segments. In [7], the authors implemented the transfer learning approach for image classification, which reduced the required amount of training data and enhanced model generalization. This approach was found

to be more efficient than cross-domain transfer learning, even with less data. For identifying the infected region, they used YOLOv8 (for image segmentation). It also demonstrated the first use of the Gazi Brains 2020 dataset, which addresses the lack of labeled and qualified brain MRI data in the medical field. Here, all experiments are implemented in PyTorch 1.12.1. According to the authors in [8], precise brain tumor segmentation from MRI images can be achieved by combining transfer learning with deep learning-based convolutional neural networks (CNNs), notably Inception-V3, VGG16, and VGG-19. Their work offers a promising method for early brain tumor identification by showing that transfer learning improves accuracy, shortens training times, and works well with sparse data annotations. In [9] a comprehensive survey of the unique security vulnerabilities exposed by the FL ecosystem, highlighting vulnerability sources and their unique challenges is provided. As the FL paradigm is getting popular day by day, many of the companies who are using this technique may not be aware of these vulnerabilities. So, their study aims to provide new perspectives and bring attention to building secure and robust FL environments suited for largescale adoption, and their work mainly touches on the issues related to FL security and not privacy. In [10] the authors pointed out that decentralized learning has several benefits, especially for applications in the medical field. They provide recommendations for secure implementation after identifying key challenges with federated learning for medical data, such as heterogeneity, client management, traceability, accountability, and security. The work in [11] suggests a unique technique to increase the accuracy of brain tumor identification from MRI images: BKNN, which combines Bagging Ensemble with K-Nearest Neighbours. The approach shows remarkable effectiveness in simplifying medical picture segmentation, achieving 97.7% categorization accuracy. Using MRI data, in [12] the authors developed a model that uses distributed federated learning and deep learning to classify brain tumors. Using cross-validation, the model obtained high accuracies of 0.82 and 0.96 on the BT-small-2c and BT-large-3c datasets, respectively. Their method avoids the requirement for a central database by ensuring data confidentiality and privacy through distributed learning among devices. Authors in [13] highlight developments as well as obstacles in systems, privacy, and statistics when talking about federated learning in healthcare informatics. They place a strong emphasis on model accuracy, expert knowledge, incentive systems, and data quality. To improve data privacy, the architecture makes use of safe federated learning, global sensitivity estimations, and local differential privacy. The incapacity of FedAvg to manage skewed data distributions, which can impair CNN performance owing to weight divergence, is a disadvantage. By contrasting FedAVG with Bayesian and non-Bayesian approaches and introducing uncertainty into predictive aggregation, authors in [14] achieved comparable results in non-distributed environments. Discussed are privacy issues and dispersed data challenges in machine learning, with a focus on larger neural networks and more communication rounds in FedAVG. Although computationally costly, Bayesian techniques provide more accurate prediction distributions, emphasizing the absence of predictive uncertainty in federated learning. The importance of precise early-stage brain tumor classification is emphasized in [15]. Concerning brain tumor identification on MRI images, authors suggest a CNN-LSTM hybrid model that achieves

noteworthy metrics for accuracy, precision, recall, and F1-measure. With the use of methods like data preprocessing, threshold, extreme point computation, and bicubic interpolation, their CNN-LSTM model successfully classifies brain tumors with an astounding 99.1% accuracy.

During the research, it is found that many of the healthcare departments are still using traditional machine learning and deep learning techniques. They are facing problems in collecting or gathering data to train the model. Hence, the necessity to use the federated learning approach to solve this problem by enhancing data privacy is decided. However, some systems that have adapted to this approach are facing some challenges and vulnerabilities during implementation. In this study, a Python framework is chosen to solve this problem. Consequently, the model is implemented using Python's FLOWER Framework [16]. It is open-source, easy to use, and flexible. It has good ML and model support. It also works with PyTorch, Tensorflow, NumPy, Transformers, MXNet, JAX, sci-kit-learn, fast ai, and Pandas and supports all major platforms like iOS, Android, and plain C++.

### 3 Proposed Methodology

For the image classification, both traditional machine learning and deep learning approaches can be followed. However, using traditional machine-learning techniques for classifying images, they often fail to identify the subtle variations and complex patterns found in large datasets, necessitating the human extraction of features from training images. To train a model using distributed data an approach called federated learning can be followed which was first introduced by Google back in 2016. The main aim of the proposed approach is to train the model that should learn the patterns from the already trained smaller models.

The first step is to gather eligible clients. Client selection needs to be performed very carefully. If this process is not done properly, then problems may arise like Model poisoning, Back Door attacks, and Evasion Attacks. Establishing secure communication between servers and clients is another big challenge. Without proper communication between clients and servers, it is difficult to make sure the trained local models are coming from real clients or not. If communication channels are not safe, some attackers can easily attack and change the weights of locally trained models or global models.

To solve this problem, a proper framework needs to be selected in the second step that makes the work easier. There are many frameworks available in the market like TensorFlow-federated, Fed ML, PySwift, and Flower, for the same task. Selecting the framework was challenging, but after analyzing, it was found that the Flower framework is the most suitable for the work. The Flower framework supports all the existing frameworks like TensorFlow, PyTorch, and transformers. Hence, the client can use any framework they want to train their local models at their ends, which may not be supported by other frameworks like TensorFlow-federated, where only TensorFlow must be used at both sides. The existing frameworks also don't have enough

model support as compared to the flower framework. Flower framework uses gRPC channels for communication, which is more secure than the traditional approach as it uses HTTP version 2. This HTTP 2 allows us to communicate using Protocol Buffers (PB), which makes this communication secure. This protocol communicates in fully duplex mode, so both clients and servers can communicate at the same time. This makes communication faster and even more secure.

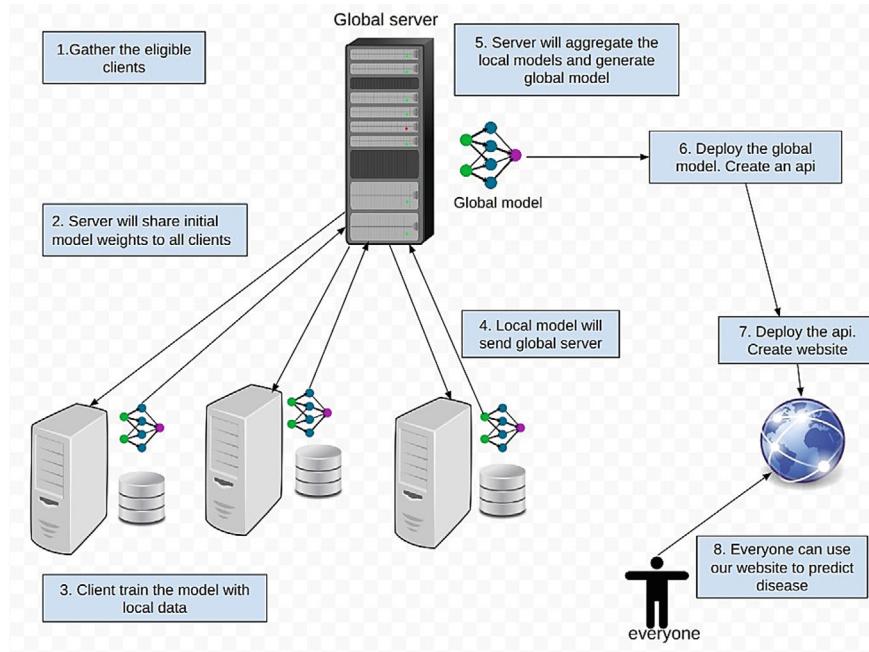
This framework contains many in-build stages and allows us to design the custom aggregation stage. Using this feature, the server-side strategy is defined to aggregate all models, which are trained on the client side. One of the most popular stages is FedAvg, which is commonly used to aggregate models at the server.

To test the system, two different clients have been chosen. This system is trained for three classes, they are glioma, meningioma, and no tumor. Client-1 has 80% glioma, 50% no tumor, and 20% meningioma in total training images. Similarly, client 2 has 20% glioma, 50% no tumor, and 80% meningioma in total training images. It can be easily seen there is an imbalance of the dataset of classes 1 and 3 in the clients. If they trained their models individually, client1 cannot predict the [class3] disease properly as it has only less data when compared to other classes in the dataset. Similarly, client2 also cannot predict [class1] as it has less data in its dataset when compared with other classes. They cannot share their hospital data for this training process. In this scenario, implementing the federated learning technique can be considered for better results and can take advantage of both clients to create a global model.

As the Flower framework for implementing federated learning is used, there arises less concern about the client on what approach they are using to train their local model as Flower can support all deep learning frameworks that are currently available in the market. So, during the training, one of the clients is using the pre-trained model VGG16 and another is using the ResNet50. With this approach, good training and testing accuracy is achieved. Figure 1 depicts the process flow of the proposed federated learning approach.

After generating the global model, this model is made available for everyone to use, as Python and its framework is used to train models, Python-related framework can be used to deploy this model. Hence, a website can be created and deploy the model accordingly. But if anyone wants to use the proposed model for brain tumor classification, they must visit the website and use the model, which is again quite inconvenient to the end users. Also, deploying the model in any mobile applications is not possible, as the model is available only on the website. Consequently, it is decided to deploy a web service and allow anyone to access the services through the internet. REST API is a way to achieve the requirements by using a set of operations.

If the model is deployed in an API, then anyone can hit the API endpoint with a query image and get the prediction of that image. There are many Python web frameworks available in the market, like Django REST, Flask, Falcon, Pyramid, etc. for creating the API. However, in the proposed approach Flask is used to create this API. Flask is one of the Python libraries [17]. It is a micro web framework, which is mainly used to create micro web services in the Python ecosystem due to its flexibility and simplicity. The flask has two extensions to create API, they are Flask



**Fig. 1** Architecture of the proposed federated learning model

RESTful and Flask RESTX. The Flask RESTful has fewer features when compared to Flask RESTX. The restful does not have namespace support and documentation support also. But RESTX has both and comes with in-built swagger documentation. This in-built support of swagger documentation of RESTX made the work of API testing and API documenting simple. Simultaneously, a web application and a mobile application is also created to test the API. The website is used to gather eligible clients and any hospital can provide their services to patients online.

As we know, the importance of gathering the clients is very important in federated learning. It is a painful task to go to every hospital and know whether that hospital is eligible or not for the training. Thereafter, to simplify this process, on the website, a client module is created. Using this module, the clients can request admin to add their hospitals to participate in the training process. The admin will verify and approve access to the clients. Further, when the server is ready to aggregate and generate the global model, then the admin can send a mail regarding the training process and timings.

Figure 2 shows the client registration page which can be used by clients or hospital admins, who are interested in registering them as the clients to participate in the training process. By entering their details, the website admin will get a request in his dashboard.

In the admin dashboard, a client applications page is created. It will display all client requests. By using this, the admin can easily contact them using their phone

**REGISTER**

## Register your Hospital

- Enter all details in respective fields
- Already registered hospital can not be registered again
- If request sent successfully, please check the given mails regularly for updates
- Please keep your email, further communication will be through mail only

Enter Your Hospital Name  Enter Your Name   
 Enter Your Email  Your Phone number   
 Enter hospital address

**Apply**

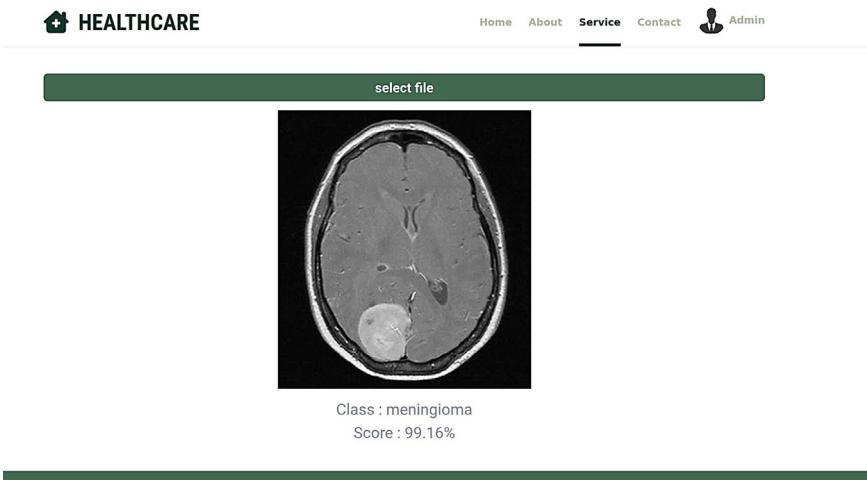
**Fig. 2** Client registration

number or email and can validate whether they are valid clients for the system or not. If they are valid, by clicking the accept button, the admin can easily permit access to the client, and the corresponding client will receive an email confirmation from the website. If the admin feels that the client is not valid to them by clicking the reject button, he can simply reject the client's application. The website admin can see all the registered clients and can contact them regarding training in the future.

Figure 3 shows the sample output or prediction of the web app. It is the main feature of the model. The end user can access this page to upload the infected brain tumor MRI image and get the classification result. This system will also give the confidence score of the predicted class along with the type of brain tumor. So that the patient or the end user, who has requested that image, can take the proper action accordingly. Figure 4 represents the screenshot of the Health Care API, which is created using Python Flask-RESTX. As discussed before, the framework itself comes with swagger integration, which makes the API testing simpler. As this is an API (a web service) anyone can hit the endpoint and get the prediction. Whenever the model is trained, it is quite difficult to delete the old model and change it for the new model. And they also need to change the code, like updating the model path with a new model path. So, to solve this issue, another endpoint is created in the API to upload the new model by overriding the old model. And everyone can access this endpoint, as the users must not have access to upload the model and to change the old model. To address this issue, this endpoint is protected. The JWT authentication is implemented, so the admin with its credentials can generate the JWT access token and can be authenticated to upload the new model.

Figure 5 refers to the interface of the web API. Here, it can be observed that the route “upload/brain-tumor-model” is protected with the JWT authorization header, as it must not be accessed by every user. The small lock icon at the end of that route can be observed. Only the admin can generate the JWT access token using the “auth/authenticate” endpoint by entering their credentials. So only the admins can access this endpoint.

Not only these services, but using the website, any hospital can easily provide their services to users online. The website has main features like appointments, paying

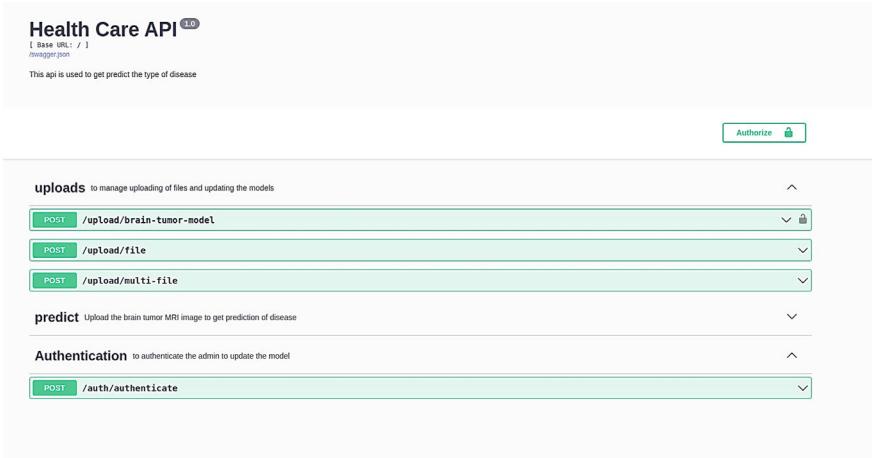


**Fig. 3** Sample output (prediction) of web app

The screenshot shows the "Health Care API" interface using the Swagger tool. At the top, it says "Health Care API 1.0" and provides the base URL: "http://127.0.0.1:5000/swagger.json". A note states: "This api is used to get predict the type of disease". Below this, there is a "predict" section with the sub-instruction: "Upload the brain tumor MRI image to get prediction of disease". A "POST /predict/brain\_tumor" button is shown. The "Parameters" section includes a "Name" column and a "Description" column. An entry for "image" is listed as required, with a "Browse..." button and the value "meningioma.jpg". There is also a "Cancel" button. Below this, there is a large blue "Execute" button. The "Responses" section shows a single entry for "200 Success". The "Response content type" is set to "application/json".

**Fig. 4** The health care web API

invoices, managing doctors and patients. The website is secure, flexible and easily scalable. For managing the database java (servlet and JSP) for back-end, HTML, CSS, JS, and AngularJS for frontend, MySQL database with hibernate framework is used. To store the medical report, Firebase (cloud storage) is used and encrypted them before storing them in the cloud.



**Fig. 5** The health care API interface

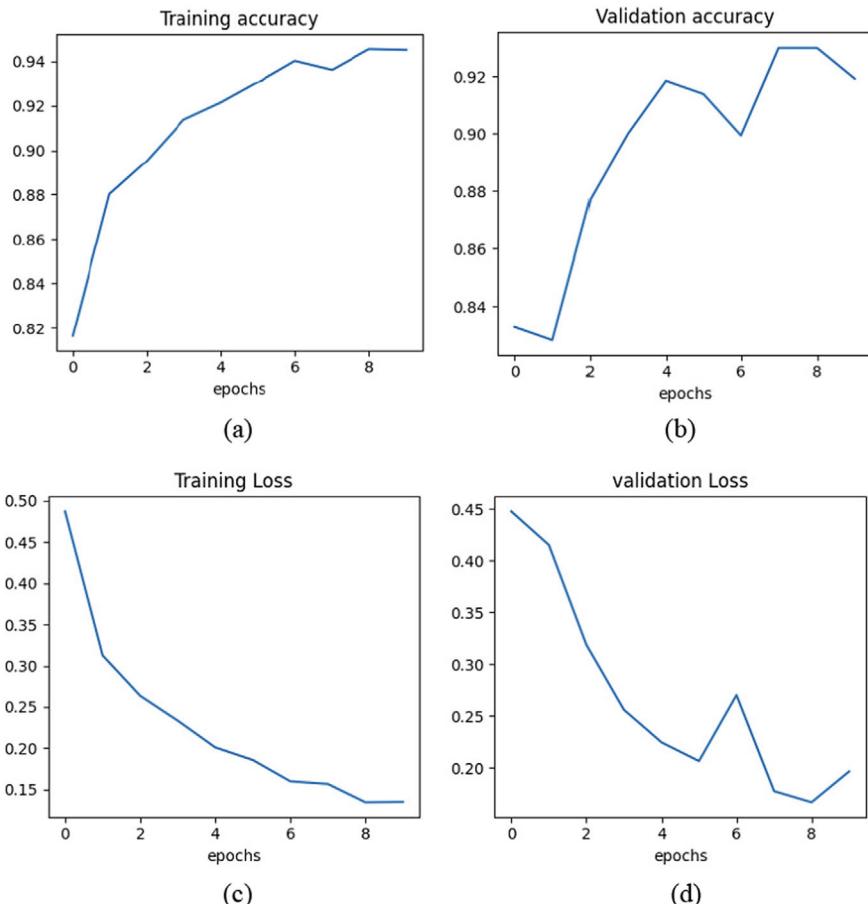
## 4 Results and Discussion

This federated learning approach has many advantages when compared to the existing traditional deep learning approach. Using this technique, data privacy can be significantly enhanced, and improve the model accuracy and diversity, and it has greater scalability capacity. These are all features that are more important features of a healthcare department if they want to train a model for decentralized data, and that can be achieved by this framework easily.

To test the system, two clients are selected who have imbalanced data. When they train individually. Then one of the clients achieved 95% training accuracy and 92% validation accuracy with their local data. Similarly, another client achieved 92% training accuracy and 90% validation accuracy with their local data. However, when these models were evaluated on the global test data, they didn't perform well as both the clients had imbalanced data in their training dataset. They achieved only 73% and 70% validation accuracy respectively on the global test dataset. Figure 6 shows the training results of one of the clients of the system. It can be observed that the training accuracy is very high, but the validation accuracy is quite low because of this imbalanced dataset. These clients can't exchange their data due to data security reasons.

However, when the federated learning approach is used to train these clients, a satisfactory result is achieved than the previous approach. Using this method, resulted in enhancing data privacy between the clients and improved the model's accuracy and diversity. The global model of the system achieved a validation accuracy of 91% on the global validation dataset.

Federated learning in the proposed system is efficiently implemented using the Flower framework. There are many advantages of using the framework when it is



**Fig. 6** Training results of one of the clients **a** Training accuracy, **b** Validation accuracy, **c** Training loss, **d** Validation loss for 10 epochs

compared to other frameworks. It has a wide range of model support; it is compatible with almost all deep learning frameworks that are available right now. So, the clients can use different frameworks to train their local model and we can aggregate them and create a global model using this framework. It also uses the gRPC protocol from communication during the training process, like the server sharing the initial training weight, and a client sending their model to the server for generating a global model. It is more secure and faster when compared to traditional HTTP RESTful communication between client and server.

A flexible API is also being developed to deploy the global model, so anyone from the internet can easily send a post request to the endpoint with the query image and get the predicted disease name easily along with the confidence score. To gather clients, a client module is incorporated in the web app. Thereafter, to train the model, searching

for clients is not necessary; they register through the web app and can participate in the training. This feature made the work easier and reduced the amount of time spent searching for clients to train in the global model.

## 5 Conclusion and Future Directions

Through this approach, the federated learning methodology is successfully realized. It resulted in achieving data privacy between the clients and improved the model's accuracy and diversity. Here only brain tumor disease is considered for the model training, but in the future, this innovative technique can be applied to other potential problems too, like skin cancer detection, diabetes detection, and so on. The FedAvg aggregation strategy is used on the server side, which is provided by the Flower framework. Considering future scope, a custom strategy can be framed and aggregate the client model in a better way to produce better output. In the proposed system, centralized relational databases (MySQL) is used to store the client's and patient's data. This is not secure, and if this database gets overloaded and goes down, then data accessibility becomes a challenge. Also, blockchain can be introduced into the system to store these details, which is more secure than this centralized relational database. Hence, whenever there is a need to store new data, a new block with that data can be created and add the block to the blockchain.

## References

1. Brain tumor: Statistics, <https://www.cancer.net/cancer-types/braintumor/statistics>. Accessed 05 Apr 2024
2. Kondaveeti HK, Sanjay KS, Shyam K, Aniruth R, Gopi SC, Kumar SV (2023) Transfer learning for bird species identification. In: 2023 2nd international conference on computational systems and communication (ICCS). IEEE
3. Abdar M, Fahami MA, Chakrabarti S, Khosravi A, Pławiak P, Acharya UR, Tadeusiewicz R, Nahavandi S (2021) Barf: a new direct and cross-based binary residual feature fusion with uncertainty-aware module for medical image classification. Inf Sci 577:353–378
4. Amran GA, Alsharam MS, Blajam AOA, Hasan AA, Alfaifi MY, Amran MH, Gumaei A, Eldin SM (2022) Brain tumor classification and detection using hybrid deep tumor network. Electronics 11(21)
5. Amin J, Anjum MA, Sharif M, Jabeen S, Kadry S, Moreno P et al (2022) A new model for brain tumor detection using ensemble transfer learning and quantum variational classifier. Comput Intell Neurosci 2022
6. Kapusiz B, Yusuf U, Koçer S, Dundar O (2023) Brain tumor detection and brain tumor area calculation with matlab. J Sci Rep-A 052:352–364
7. Terzi DS, Azginoglu N (2023) In-domain transfer learning strategy for tumor detection on brain MRI. Diagnostics 13(12):2110
8. Alla S, Athota K (2022) Brain tumor detection using transfer learning in deep learning. Indian J Sci Technol 15(40):2093–2102
9. Bouacida N, Mohapatra P (2021) Vulnerabilities in federated learning. IEEE Access 9:63229–63249

10. Yoo JH, Jeong H, Lee J, Chung T-M (2022) Open problems in medical federated learning. *Int J Web Inf Syst* 18(2/3):77–99
11. Archana KV, Komarasamy G (2023) A novel deep learning-based brain tumor detection using the bagging ensemble with k-nearest neighbor. *J Intell Syst* 32(1):20220206
12. Mahlool DH, Abed MH (2022) Distributed brain tumor diagnosis using a federated learning environment. *Bull Electr Eng Inform* 11(6):3313–3321
13. Xu J, Glicksberg BS, Su C, Walker P, Bian J, Wang F (2021) Federated learning for healthcare informatics. *J Healthc Inform Res* 5:1–19. <https://doi.org/10.1007/s41666-020-00082-4>
14. Thorgeirsson AT, Gauterin F (2020) Probabilistic predictions with federated learning. *Entropy* 23(1):41. <https://doi.org/10.3390/e23010041>
15. Alsubai S, Khan HU, Alqahtani A, Sha M, Abbas S, Mohammad UG (2022) Ensemble deep learning for brain tumor detection. *Front Comput Neurosci* 16:1005617. <https://doi.org/10.3389/fncom.2022.1005617>
16. Beutel DJ, Topal T, Mathur A, Qiu X, Fernandez-Marques J, Gao Y, Sani L, Li KH, Parcollet T, de Gusmão PPB et al (2020) Flower: a friendly federated learning research framework. <https://doi.org/10.48550/arXiv.2007.14390>
17. Python, <https://weexperto.com/technology/python/>. Accessed 05 Apr 2024

# **Computer Networks**

# Low Cost Emergency Communication System for Disaster Affected Areas



Sanjoy Debnath, Yaddanapudi Srilekha, Vibhuthi Amarnath,  
and Yaddanapudi Venkata Sri Harsha

**Abstract** During natural disasters, breakdowns in the infrastructure based communication systems often hinder emergency response efforts, leading to loss of life. To address this challenge, we propose an emergency communication system utilizing LoRa modules, capable of operating without cellular towers or base stations. Our system aims to provide fast and reliable communication in challenging environments where infrastructure is compromised. By efficiently connecting victims with emergency management organizations, the system increases the likelihood of saving lives and facilitates faster rescue operations. Users can connect to this network to exchange information and access vital services. The system also includes emergency services integration, enabling coordination and data dissemination during critical moments. Key features of the system include real-time communication, data sharing, and location tracking, enhancing situational awareness and aiding in rescue efforts. Overall, the designed emergency communication system offers a comprehensive solution to address communication challenges during disasters. Its adaptability and resilience make it suitable for various emergency scenarios, ranging from natural disasters to large-scale accidents. By providing a robust communication infrastructure, the system plays a crucial role in improving emergency response capabilities and ultimately saving lives in crisis situations.

**Keywords** Emergency communication system (ECS) · Low cost ECS · LoRa module · NodeMCU

## 1 Introduction

Natural disasters like floods, cyclones, and earthquakes mostly affect telecommunication services. In such situations, effective communication is paramount for coordinating emergency response efforts and saving lives [1]. However, traditional communication infrastructure often fails during such crises, leaving victims stranded and

---

S. Debnath (✉) · Y. Srilekha · V. Amarnath · Y. V. S. Harsha

Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India  
e-mail: [drsanjoydebnath@veltech.edu.in](mailto:drsanjoydebnath@veltech.edu.in)

unable to reach out for help [2]. To address this critical challenge, we propose a node microcontroller unit (NodeMCU) based emergency communication system. This innovative system leverages the versatility and affordability of microcontroller, a compact single-board computer, to establish a resilient communication network in disaster-stricken areas. By integrating long-range (LoRa) modules, the system can operate independently of cellular towers or base stations, ensuring communication continuity even in blackout zones. The NodeMCU serves as the central processing unit to create an ad-hoc network. This network enables real-time communication, data sharing, and location tracking among victims and emergency responders. Furthermore, the system integrates emergency services, facilitating coordination and resource allocation during critical moments. With its adaptability and reliability, the NodeMCU-based emergency communication system offers a lifeline in times of crisis. By providing fast and reliable communication capabilities in challenging environments, it enhances the efficiency of emergency response efforts and ultimately helps save lives.

## 2 Related Works

The Raspberry Pi is a very useful device to make smart communication systems for various applications. In [3], the authors comprehensively analyzed the Raspberry Pi-based model for small and affordable systems. The developed systems could have all the benefits of computers to solve real-life problems. Emergency communication systems based on microcontrollers like NodeMCU are very popular for their cost-effectiveness and efficiency. In [4], the authors propose an IoT-based system utilizing MEMS sensors and Arduino in vehicles to detect accidents and transmit the victim's location data to the cloud, which is then retrieved by an ambulance's NodeMCU system, aiming to reduce death rates by providing timely medical assistance. The developed communication system utilizes sensors and actuators as input, enabling applications like Thingspeak and Raniso in mobile for the needful emergency services at the required moment. Based on the same NodeMCU microcontroller, an emergency communication system (ECS) is developed utilizing the ThingSpeak platform [5]. The developed ECS is able to provide needful rescue and telecom services to remote users at the time of need. In the designed model, transmission of data signals is done by the LoRa module. In [6], the author proposes LoRa preamble detection, deriving optimal detectors and designing pREAMbles to maximize detection probability while maintaining a target false alarm rate, revealing suboptimal methods, and providing engineering insights, ultimately proposing a scheme outperforming existing ones by over 10 dB in SNR. WiFi plays a vital role in telecommunications and data transmission. In [7], the author proposes an optimized model integrating WiFi and wireless mesh networks for cost-effective, coverage-efficient, and capacity-maximizing network design, catering to emergency services and environmental monitoring in areas lacking telecommunication services, utilizing linear programming for dimensioning. The designated system, named high availability and disaster (HADR) is used

in ECS [8]. It introduces AASAPS-HADR, an integrated system for autonomously surveying and planning emergency responses in disaster areas, focusing on situational awareness through autonomous aerial surveys and proposing AI algorithms for perception and planning, along with initial experimental results for non-contact health monitoring, aiming to enhance disaster response capabilities [8]. In the domain of underwater acoustic communication, the author utilizes biological parameters underwater, integrating temperature and heart rate sensors with an LCD display for data visualization. In the proposed system, the author also integrates GPS-enabled emergency alerts for immediate rescue. To monitor divers' biological conditions underwater, the author employs temperature and heart rate sensors. The resulting data is displayed on an LCD, enabling emergency alerts with GPS location transmission for immediate rescue and enhancing safety during underwater research [9]. An innovative system that incorporates speech recognition, compression, coding, and transmission algorithms is proposed in [10] for real-time audio signal processing in emergency broadcast. The authors utilize Python-based speech detection techniques for noisy signals and employ formant theory for speech recognition, enabling real-time analysis of signal processing algorithms' performance. A mobile application featuring a deep learning system tailored for residence monitoring, specifically designed for ECS [11], in [11], the author introduces a mobile application integrated with a deep learning system for residence monitoring, addressing concerns over theft in Indonesia by providing features like camera capture, motion detection, SOS alerts, and notification of dangerous activities, ensuring user security and comfort. Towards the enhancement of communication between emergency management agencies and the public, the author explores improving communication between emergency management agencies and the public through an organizational analysis framework, emphasizing the need for standardized training, additional resources, and interventions from technical communicators to improve communication strategies [12]. By acknowledging the fluctuations in RSSI values induced by obstructions and distance variations, this study aims to provide a robust and reliable method for indoor positioning. In [13], the author develops post-disaster indoor position tracking using wireless sensor networks, employing NodeMCU ESP8266 to transmit RSSI values for estimating wrist strap wearer positions and noting variations in RSSI values based on obstructions and distance. By seamlessly integrating these components, the framework aims to enhance communication and coordination among emergency services. In [14], the author introduces an end-to-end framework utilizing ad-hoc mobile networks and contextual intelligence to address technology constraints, demand-supply management, and information access issues in disaster response, enhancing communication and coordination among emergency services. Through meticulous evaluation, this approach promises to advance the efficiency and reliability of device-to-device (D2D) communication networks in ECS. The author introduces a centralized low-latency scheduling policy for uplink underlay D2D communication, adapting the Server Side Greedy algorithm, demonstrating its efficacy, particularly in large systems with improved performance [15]. A pioneering open-source Wi-Fi localization-based occupancy detection system prototype was specifically crafted for emergency management in smart buildings. In [16], the author

presents an open-source Wi-Fi localization-based occupancy detection system prototype tailored for emergency management in smart buildings, emphasizing real-time performance, low latency, and high accuracy, with detailed design, implementation, testing, and performance evaluation. These strategies are designed to fortify continuity and elevate the quality of healthcare services amid the chaotic backdrop. Key five strategies for healthcare organizations to enhance IT-enabled disaster response are discussed in [17], based on a case study of the Veterans Health Administration's response to Hurricane Katrina, aiming to improve continuity and quality of healthcare services during natural disasters [17]. The disaster response systems and rescue operations available in the literature are not timely and cost-efficient. The lack of an emergency communication system in literature based on cost efficiency can be a powerful narrative approach that draws attention to how vulnerable both individuals and communities are at the time of any emergency. The absence of an effective emergency system in the literature underscores the importance of developing robust, accessible technology to ensure safety and rescue at the event time, regardless of cost and performance. Thus, in this work, we designed a low-cost ECS based on the LoRa module and Node MCU for long range emergency communication. The main contributions of this work are summarized below:

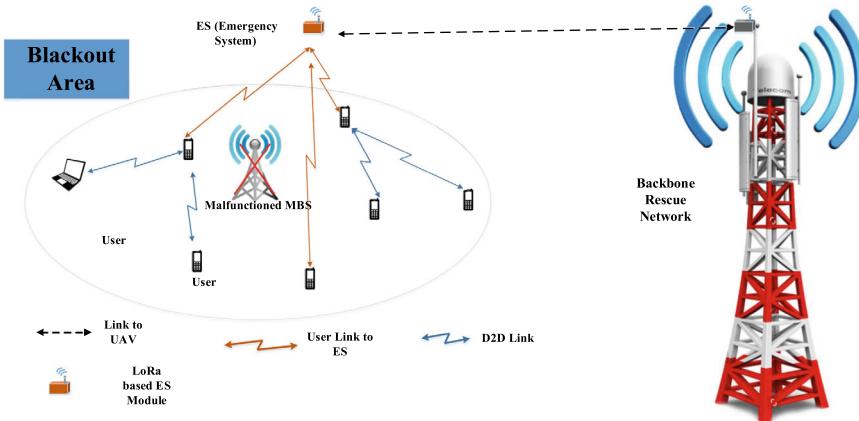
- (i) Designed a low cost emergency communication system based on LoRa module and Node MCU, to assist the mobile user in any emergency.
- (ii) Developed SOS application for the assistance in emergency.
- (iii) Test the designed transmitter and receiver in practical field to validate the results.

The rest of the sections are as follows: In Sect. 3 of the manuscript, we discuss the system model of the proposed ECS, followed by results and discussion in Sect. 4. Finally, the conclusion is drawn in Sect. 5 of the manuscript.

### 3 System Model

The system model for a NodeMCU-based emergency communication system utilizes the NodeMCU's versatile capabilities along with LoRa to ensure reliable communication during critical situations. The NodeMCU serves as the central control unit, managing communication between components and processing emergency alerts. The LoRa module enables long-range communication, ideal for transmitting data over several kilometers to provide connectivity to local networks for internet access. Sensors or manual triggers detect emergencies, prompting the system to initiate alerts. With robust data processing capabilities, the NodeMCU formats and transmits emergency information using established communication protocols. The developed system model is shown in Fig. 1.

This system model integrates multiple functionalities to create a dependable emergency communication solution capable of swift and effective response in challenging environments. As shown in Fig. 1, the designed ECS based on LoRa and NodeMCU



**Fig. 1** Developed LoRa based emergency system

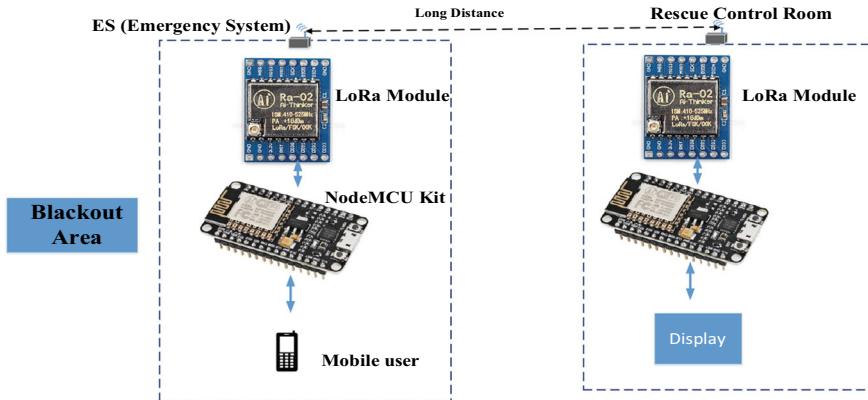
providing the requisite services in the blackout area at the event time. After a disaster, a large geo-graphical area can come under blackout; at that time, efficient ECS can be established for the rescue operation. As shown in Fig. 1, the proposed ECS system establishes communication with the rescue operation team, which is situated at a very large distance from the event location, through LoRa module. The parameters assumed for the LoRa module are described below in Table 1.

The circuit of the proposed ECS is given below in Fig. 2. As shown, the ES is situated in the event location, and the disaster-affected users of the event location will get the WiFi service through the NodeMCU for the emergency service. In emergencies, information is transferred to the rescue operation team, situated at a very long distance. This long-distance communication is established through LoRa module.

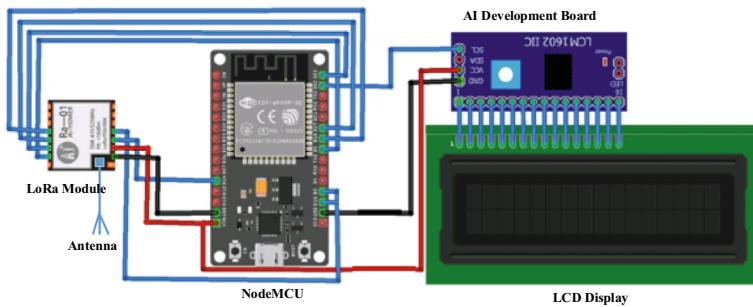
The proposed ECS is designed using NodeMCU and a LoRa module. In designing an ECS, pin layout coordination is crucial for the seamless integration of components. Utilizing a LoRa module, NodeMCU, LCD display, and Arduino development board, pin allocation should prioritize efficient data exchange and functionality. Assigning appropriate pins for data transmission and reception between the LoRa module and NodeMCU, as shown in Fig. 3, will ensure reliable long-range communication capabilities. The proposed system utilizes both hardware and software platforms. The general process of making the system is discussed below:

**Table 1** Parameter assumption for LoRa module

Parameters	Values
Frequency of operation	863–870 and 902–928 MHz
Coverage distance	15 km
Power consumption	Lower than Zigbee
Data rate	10 kbps



**Fig. 2** Layout of the developed ECS



**Fig. 3** Pin layout diagram of the developed model

Meanwhile, connecting the LCD display to display critical information demands careful consideration of pin compatibility and functionality. Through meticulous pin layout planning, this emergency communication system can effectively transmit vital data and facilitate rapid response during crises.

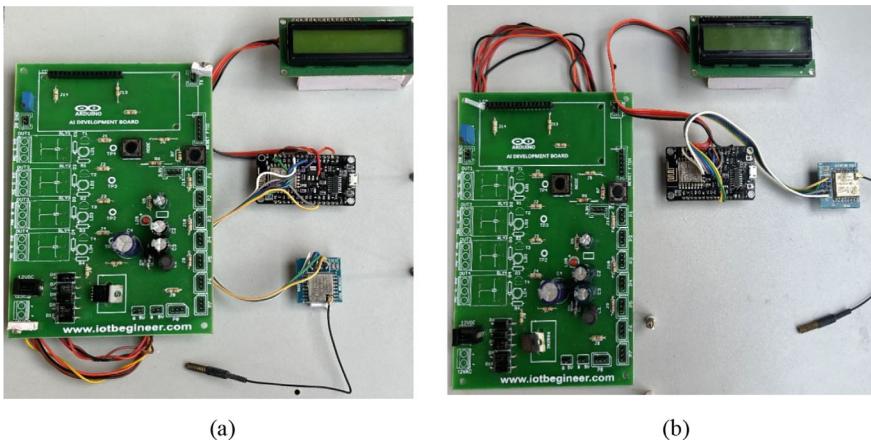
The first step consists of installing the requisite software on the NodeMCU module. For which we connect the NodeMCU to the internet via Ethernet and flash the NodeMCU with the appropriate firmware that supports LoRa communication. We use Arduino IDE to install the software. The connection of the LoRa module to the NodeMCU will ensure the connections are correct according to the specifications of your LoRa module. The separation of this module from the NodeMCU is according to its specifications. Installation of necessary software packages such as Python, Node.js, related libraries is required for communication. The required data for a Python or Node.js script to handle communication with the NodeMCU and to send or receive emergency messages. The code of Arduino sketch to handle LoRa communication with other necessary tasks, such as sensor data collection, ensures that it is programmed to listen for incoming messages and send emergency messages

when triggered. The NodeMCU will be connected to the same network. This could involve commands for sending emergency messages, requesting status updates, etc. The different setup steps include:

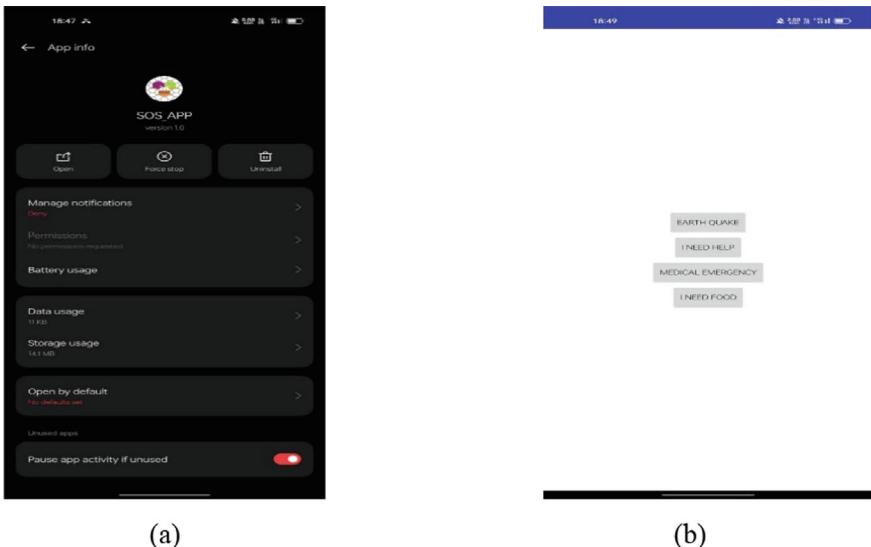
- (a) *System Design and Architecture*: Define the overall system architecture, including the roles of NodeMCU nodes, and LoRa modules. Determine the communication protocols and data formats for transmitting messages between components.
- (b) *Software Installation and Configuration*: Install the required software packages, including the operating system and communication libraries (e.g., LoRa WAN stack). Flash the NodeMCU boards with the appropriate firmware, such as the NodeMCU firmware or custom firmware for LoRa communication. Configure the NodeMCU boards to connect to the LoRa module and establish communication.
- (c) *NodeMCU Setup*: Develop firmware for the NodeMCU boards to collect sensor data, format emergency messages, and transmit them. Implement error-handling mechanisms to ensure reliable message transmission. Test the NodeMCU firmware in various scenarios to validate its performance and reliability.
- (d) *LoRa Communication Configuration*: Configure the LoRa modules on NodeMCU boards to establish communication with neighboring nodes. Define LoRa parameters such as frequency, spreading factor, and transmission power to optimize range and reliability. Implement routing algorithms to relay messages between distant nodes and backhaul.
- (e) *Testing and Validation*: Conduct comprehensive testing of the entire system under simulated emergency scenarios, including network congestion, node failures, and environmental interference. Evaluate the system's performance in terms of message delivery latency, reliability, and coverage range. Iterate on the design and implementation based on testing results and user feedback to improve system robustness and usability.
- (f) *Deployment and Maintenance*: Deploy the emergency communication system in real-world scenarios, such as disaster-prone areas or remote regions with limited connectivity. Provide training and documentation for system operators and emergency responders on system usage, troubleshooting, and maintenance. Establish procedures for regular maintenance, software updates, and expansion of the communication network as needed.

The developed ECS model consists of two transceivers, one for the event location and another for the rescue control room, as shown in Fig. 4a and b respectively. Figure 4a depicts the transceiver of the ECS for the event location. The transceiver intended for the rescue control room is shown in Fig. 4b.

To access the emergency service, the developed SOS application software must be installed on the mobile device for necessary interaction with the developed ECS. Figure 5a shows the application information, while Fig. 5b depicts the interface of the developed application.



**Fig. 4** Developed hardware model for the ECS. **a** Transceiver for the event location. **b** Transceiver for the rescue control room



**Fig. 5** Developed SOS application. **a** Information about the application. **b** Interface of the developed application

## 4 Results and Discussion

Here we analyze the three most important parameters for any emergency communication system, namely, distance, latency, and throughput. Evaluating distance, latency, and throughput in communication is paramount for ensuring efficient and

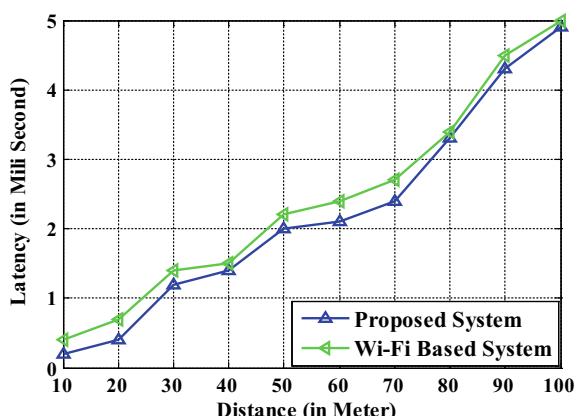
reliable information exchange across various platforms and networks. First, we analyze latency in communication vs. distance between the transmitter and receiver. Secondly, we analyze the achieved throughput based on the distance of the receiver from the ECS. Finally, analyze the latency vs. achieved throughput characteristics of the developed ECS.

Latency measures the delay between sending and receiving data packets, which is crucial for real-time applications like video conferencing and online gaming. Figure 6 depicts the relationship between latency and distance in emergency communication. As distance increases, latency tends to rise, potentially affecting response times, which are crucial in emergency situations. The data suggests a clear trend of latency amplifying with greater distances, posing challenges for effective communication during critical moments.

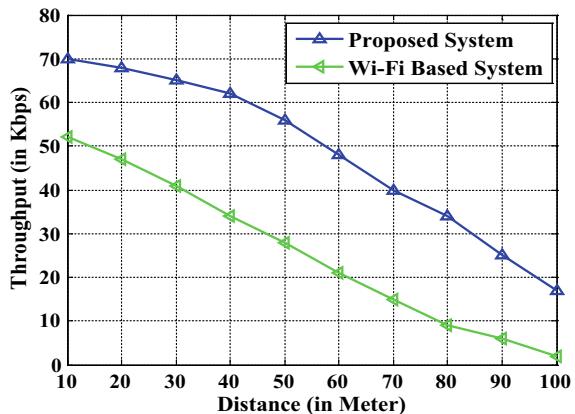
Throughput quantifies the amount of data transferred over a period, indicating network capacity and bandwidth availability. Figure 7 illustrates the relationship between distance and throughput in emergency communication scenarios. The data suggests that as the distance between communication endpoints increases, throughput tends to decrease. This decrease in throughput over longer distances could potentially hinder the transmission of critical information during emergency situations, impacting the efficiency of response efforts.

Figure 8 illustrates the correlation between latency and throughput in emergency communication settings. The data indicates that as latency increases, throughput tends to decrease, suggesting a trade-off between communication speed and data transmission capacity. This trade-off could have significant implications for emergency response efforts, as delays in transmitting information may hinder critical decision-making processes. The relationship between latency and throughput is visually depicted, underscoring the importance of balancing these factors to ensure optimal performance in emergency communication systems.

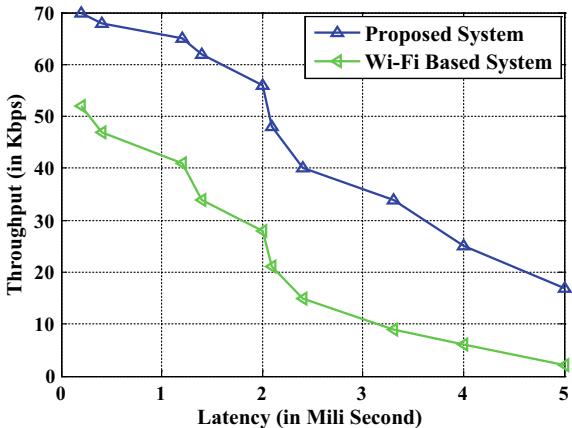
**Fig. 6** Latency in communication versus Distance



**Fig. 7** Achieved throughput based on the distance of receiver



**Fig. 8** Latency versus achieved throughput



## 5 Conclusion

Communication plays a vital role in our day-to-day life. When natural disasters or calamities occur they drastically affect the contemporary networks of communication, which has a major impact on life and death matters. Thus to avoid such cases we have proposed an alternative method of communication using NodeMCU and LoRa transceiver which will offer service at the time of emergency when other communication services are unavailable. In designing of the ECS our main motive is to make a cost effective solution for the normal mobile user. We have chosen LoRa as our transceiver module because when compared to other potential transceiver modules LoRa transceiver can work at low power for long ranges and at preferable data rates. We have used the NodeMCU in the developed emergency system in both side because it is efficient and able to handle more complex operations. Practical validation

of the developed model ensures the seamless connectivity in the emergency situation. The efficiency of the developed ECS model is confirmed by the performance-evaluating metrics pertaining to the latency corresponding to the throughput parameter and distance, as well as the analysis of achieved throughput versus distance. The developed model also ensures a reliable connectivity to communicate during any emergency situations like natural disasters or calamities.

## References

1. Debnath S, Arif W, Roy S, Baishya S, Sen D (2021) A comprehensive survey of emergency communication network and management. *Wirel Pers Commun* 124:1375–1421
2. Debnath S, Baishya S, Sen D, Arif W (2024) LTE cell planning for resource allocation in disaster communication: an Indian perspective. *Wirel Pers Commun* 135:1035–1076
3. Karthikeyan S, Raj RA, Cruz MV, Chen L, Vishal JLA, Rohith VS (2023) A systematic analysis on raspberry Pi prototyping: uses, challenges, benefits, and drawbacks. *IEEE Internet Things J* 10(16):14397–14417
4. Karthikeyan S, Kiruthik J, Madumitha S, Manikandan R, Prakash Raj V (2023) Design and implementation of IoT based accident detection and prevention system. In: *L2023 7th International Conference on Trends in Electronics and Informatics (ICOEL)*, pp 412–419, Tirunelveli, India
5. Djehaiche R, Aidel S, Sawalmeh A, Saeed N, Alenezi AH (2023) Adaptive control of IoT/M2M devices in smart buildings using heterogeneous wireless networks. *IEEE Sens J* 23(7):7836–7849
6. Kang J-M, Lim D-W, Kang K-M (2022) On the LoRa modulation for IoT: optimal preamble detection and its performance analysis. *IEEE Internet Things J* 9(7):4973–4986
7. Michael R-A, Carlos ZP, Juan I-O (2017) Optimum design and dimensioning model of a Mesh-WiFi network for emergency services in protected areas. In: *2017 IEEE second Ecuador technical chapters meeting (ETCM)*, pp 1–6. Salinas, Ecuador
8. Allen R, Mazumder M (2020) Toward an autonomous aerial survey and planning system for humanitarian aid and disaster response. In: *2020 IEEE aerospace conference*, pp 1–11, Big Sky, MT, USA
9. Vivekanand CV, Inbamalar TM, Kavimani B, Inbamalar TM, Keerthi K, Krishna VM, Kavimani B (2023) Remote health monitoring of divers using underwater acoustic communication. In: *2023 international conference on recent advances in electrical, electronics, ubiquitous communication, and computational intelligence (RAEEUCCI)*, pp. 1–6, Chennai, India
10. Denis K, Anisimov P, Statsenko L (2019) Audio signal transmission over low-speed communication channels for emergency broadcast. In: *2019 international science and technology conference “EastConf”*, pp 1–4, Vladivostok, Russia
11. Putra DD, Shalaimanda W (2022) Mobile application development for monitoring dangerous situations in smart home warning system. In: *2022 8th international conference on wireless and telematics (ICWT)*, Yogyakarta, Indonesia, pp 1–5
12. Cosgrove SJ (2023) Improving technical and risk communication: an organizational study of north Carolina emergency management and hurricane Florence. *IEEE Trans Prof Commun* 66(3):284–299
13. Magwili GV, Linsangan NB, Marasigan JMC, Villanueva CJV (2021) Post disaster indoor position tracking device with pulse detection in wireless sensor networks. In: *2021 IEEE 13th international conference on humanoid, nanotechnology, information technology, communication and control, environment, and management (HNICEM)*, pp 1–6. Manila, Philippines
14. Lohokare, J. and Dani, R (2021) An Intelligent cloud ecosystem for disaster response and management leveraging opportunistic IoT mesh networks. In: *2021 international conference*

- on information and communication technologies for disaster management (ICT-DM), pp. 125–133, Hangzhou, China
- 15. Mete A (2019) Low latency scheduling for D2D communication. In: 2019 11th international conference on communication systems & networks (COMSNETS), pp 448–450, Bengaluru, India
  - 16. Khoche S, Sasirekha G, Bapat J, Das D (2020) Near real-time occupancy detection for smart building emergency management: a prototype. In: 2020 IEEE international symposium on smart electronic systems (iSES) (formerly iNiS), pp 115–120, Chennai, India
  - 17. Bala H, Venkatesh V, Venkatraman S, Bates J (2016) If the worst happens: five strategies for developing and leveraging information technology-enabled disaster response in healthcare. *IEEE J Biomed Health Inform* 20(6):1545–1551

# A Slot-Integrated Based Partial Ground and Tapered Patch Antenna for Satellite Communications



J. Josiah Samuel Raj and G. Anitha

**Abstract** In this research, a new antenna arrangement that is optimized for the requirements of communication systems operating at 28 GHz is presented. The purpose of the design is to fulfill the requirement for high-performance antennas that are suitable for use in radar and satellite applications despite their broadband capabilities. The suggested antenna arrangement is made up of a rectangular microstrip patch that has tapered corners and a slot that is strategically positioned. This patch is produced on a FR4 substrate that has a dielectric constant of 4.4 and a height of 1.6 mm. The optimization of the design is accomplished by the utilization of HFSS software and other methodologies, including inset input and partial ground plane implementation. At a frequency of 27.54 GHz, the antenna exhibits a number of outstanding performance measures, including an impedance matching of  $50.0068\ \Omega$ , a return loss of  $-80.75\text{ dB}$ , and a Voltage Standing Wave Ratio (VSWR) of 1.0002 at the same frequency. Additionally, it achieves a significant gain of 5.0 dB and demonstrates an extraordinarily broad bandwidth of 6.06 GHz, much beyond the capabilities of conventional systems in terms of flexibility and efficiency. Not only does this study work provide evidence that the proposed antenna structure is successful, but it also highlights the fact that it has the potential to accelerate the development of future communication systems that operate at millimeter-wave (mmW) frequencies.

**Keywords** Antenna configuration · Inset input · Performance metrics · Impedance matching · Return loss · VSWR · Gain · Millimeter wave

---

J. Josiah Samuel Raj · G. Anitha (✉)

Centre for Applied Research, Institute of Electronics and Communication Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Chennai, Tamil Nadu, India

e-mail: [anithag.sse@saveetha.com](mailto:anithag.sse@saveetha.com)

J. Josiah Samuel Raj

e-mail: [josiahsamuelraj9075.sse@saveetha.com](mailto:josiahsamuelraj9075.sse@saveetha.com)

## 1 Introduction

The exponential growth of the high-performance, ultra-fast mobile device market is driven by the insatiable desire for instantaneous information access in contemporary society [1]. In the midst of the ongoing technological revolution, RADAR technology is positioned to bring about a significant transformation in communication infrastructure by offering unprecedented improvements in capacity, connectivity, and data transfer rates. It is anticipated that forthcoming satellite applications will make use of millimeter-wave frequencies, including but not limited to 28, 38, and 60 GHz [2]. However, in light of the present viability of lesser frequencies such as 27 and 37 GHz, researchers are also investigating their potential. Amidst this ever-changing environment, the design of antennas assumes a critical role in mobile communication systems, requiring the formidable task of striking a balance between lightweight construction, compactness, and optimal performance [3].

Microstrip patch antennas, which are well renowned for their streamlined look and simple production procedures, have become leading contenders in the creation of next-generation communication devices. This is because of the fact that they are very easy to manufacture. The versatility and efficiency of these antenna designs are superior to those of traditional antenna designs [4].

The pervasive impact of wireless communication on everyday lives has generated an unparalleled increase in data requirements and network speeds, forcing a fundamental transformation in network infrastructure and antenna development to accommodate the expanding needs of contemporary communication systems [5].

Advancements in satellite and radar applications continue to increase the demand for high-performance antennas in millimeter-wave communication systems, specifically at 28 GHz. This research endeavors to address this need by introducing an innovative antenna design that has been carefully customized to fulfil the rigorous specifications of contemporary communication infrastructures. The antenna configuration under consideration incorporates an unprecedented combination of state-of-the-art characteristics, such as a slot strategically placed and a rectangular microstrip patch featuring intricately tapered corners [6]. The implications of these results transcend mere technical accomplishments; they emphasize the revolutionary capacity of the suggested antenna design in influencing the domain of forthcoming communication systems [7].

Microstrip patch antennas typically have narrow bandwidths, which can limit their suitability for mmW frequency applications. Also Achieving high radiation efficiency, especially at higher frequencies, can be challenging due to various factors such as substrate losses and conductor losses. A thicker substrate reduces the substrate's influence on the resonant frequency and allows for wider bandwidth. Tapering the edges of the patch antenna can reduce edge effects and improve impedance matching, thus increasing bandwidth. By conducting thorough simulations and meticulous analysis, the proposed antenna design demonstrates exceptional performance in critical performance metrics.

## 2 Related Study

In the domain of communication systems that we are currently operating in, it is hard to overstate the significance of low-profile antennas that exhibit exceptional performance throughout a broad spectrum of frequencies with regard to their value. Among the several possibilities that were taken into consideration, MSPAs are particularly tempting due to the fact that they are cost-effective, portable, and have manufacturing methods that are not overly difficult [8]. The versatility of these components in terms of polarization, resonant frequency, radiation pattern, and input impedance further strengthens their appeal for a broad variety of applications.

Due to their weak directivity, restricted gain, radiation inefficiency, and limiting bandwidth, MSPAs are not appropriate for future satellite communications [9]. This is despite the fact that they have these intrinsic benefits. For the purpose of overcoming these challenges, a wide range of design techniques have been researched and documented. For the purpose of enhancing radiation efficiency and bandwidth, researchers have investigated a wide range of methods, including serial feeding, multi-layer substrates, multi-patch topologies, slot integration, defective ground planes, and dimensional optimization [10]. In addition, the use of approaches such as the exploitation of microstrip feed-lines as inset-feeds and partial ground planes has been introduced.

Furthermore, insufficient impedance matching at interfaces continues to be a serious problem [11], despite the fact that these methods typically generate positive consequences, such as reduced return losses and greater bandwidth at resonant frequencies. Because of the difference that was discussed before, there is a possibility that total return losses will be significantly increased, which would in turn impede the efficiency of wireless communication systems [12]. Despite the fact that various designs have shown to be effective in specific areas, such as beam-gain or reduced VSWR, they frequently fail to produce a complete increase in all essential performance measures, particularly bandwidth, which is of the biggest importance for satellite technology applications [13]. Various parameters can significantly impact the characteristics of microstrip antennas. Higher dielectric constants typically result in a lower resonant frequency and narrower bandwidth. Thicker substrates generally lead to lower resonant frequencies and broader bandwidths. Larger patches typically resonate at lower frequencies and may exhibit wider bandwidths. Proper positioning of the feed point is crucial for achieving desired performance. Introducing slots or apertures in the patch or ground plane can alter the distribution of currents and capacitance, effectively lowering the resonant frequency of the antenna. Careful design of slot geometry is essential for achieving desired performance. Changing the location of the feed point on the patch can also impact the resonant frequency. Moving the feed point towards the edges of the patch can decrease the resonant frequency, while moving it towards the center can increase it.

This research aims to create and assess a rectangular single-element MSPA that operates at a frequency of 28 GHz and is built particularly for use in satellite-based infrastructure systems by utilizing this technology. The research will be carried out

in order to accomplish this objective. This investigation is being carried out with the purpose of ensuring that a gap in knowledge that has been recognized is filled. The major objective is to improve critical performance metrics such as gain, bandwidth, and radiation efficiency. Some of the strategic methods that will be utilized in order to accomplish this objective include inset-feed, partial ground plane for impedance matching, and tedious dimension tuning via patch corner tapering.

### 3 Methodology

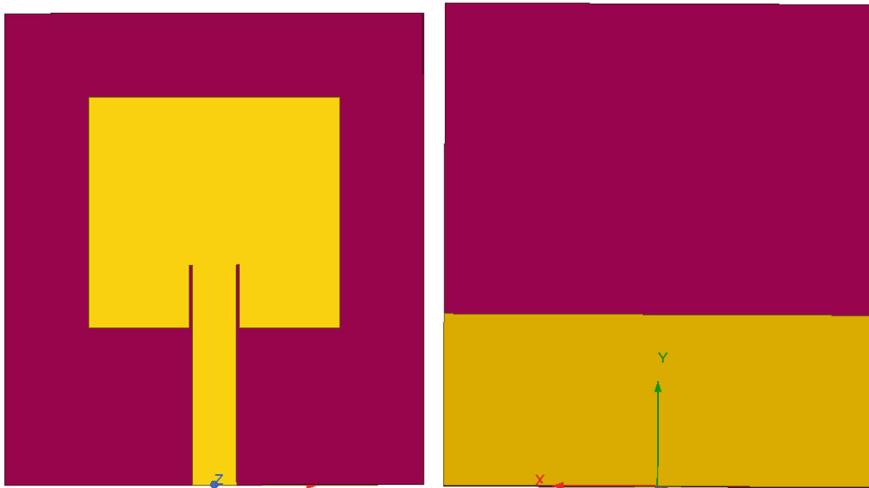
The characteristics of antenna such as shapes, dimensions, and materials are among the most essential aspects that have an effect on the design and performance of the product. This research makes use of the rectangular microstrip patch antenna because it has a broader bandwidth, is simple to construct and analyze, and is relatively simple in comparison to other antenna designs. The enhancement of antenna performance is predicated on attaining particular attributes at a specified frequency. The design process for rectangular MSPAs generally commences with pivotal determinations concerning the thickness, composition, and operational frequency of the substrate. The antenna demonstrates an array of outstanding characteristics when it comes to millimeter-wave communication systems [14].

Both Figs. 1 and 2 offer a visual representation of the MSPA, which is the focus of the analysis that is being carried out in this particular study. A FR-4 substrate with a height of 1.6 mm, a tangent loss of 0.0025, and a dielectric constant of 4.4 is used in the construction of this device. And in addition to that, it is constructed with a layer of radiating copper. In particular, it is created with the intention of functioning at a frequency of 28 GHz.

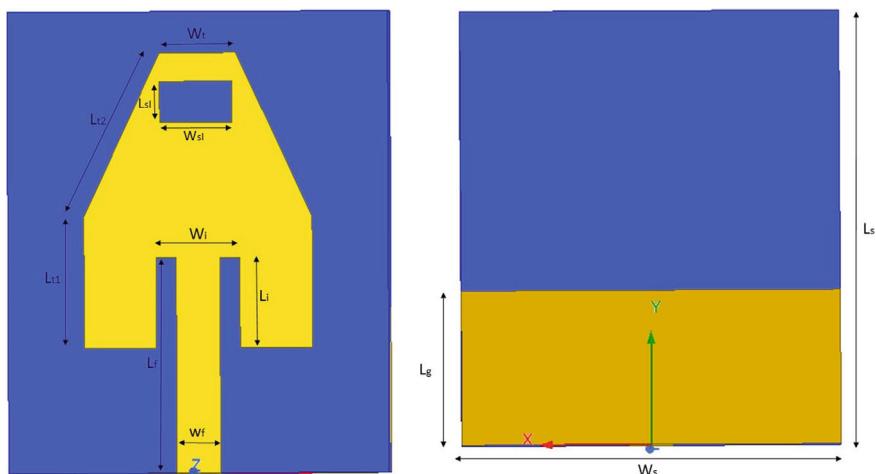
An assessment of the proposed antenna was performed utilizing Ansys HFSS 2021. It is worth mentioning that the excitation mechanism utilized was an inset feedline, which deviated from the traditional microstrip feed configurations.

The primary objective of this research is to enhance the effectiveness of a microstrip patch antenna that was designed particularly for satellite communications and operates at a frequency of 28 GHz. At first, a square patch with dimensions of 10 mm by 10 mm was partially grounded, which led to a minimum R.L of  $-19$  dB measured at a frequency of 26.4 GHz. In addition to that, it displayed a bandwidth of 4.3 GHz. After that, the tests revealed a return loss of  $-21$  dB at a frequency of 27.4 GHz and a bandwidth of 5 GHz [15].

The design of the second antenna is designed by tweaking the patch's form by shifting its width and length, as well as altering the breadth and depth of the inset gap. In addition, the patch has tapering corners at both ends, with slots positioned opposite to the port. Figure 2 vividly displays the numerous elements and creative features that set it apart from previous versions. This depiction provides a detailed understanding of the antenna's structure and shape, emphasizing the precise changes and deliberate improvements that have been carefully integrated into its design. These adjustments result in improved performance measurements, including as R.L, bandwidth, VSWR,



**Fig. 1** Initial rectangular MSPA. (Top and bottom view)



**Fig. 2** Novel slot integrated tapered MSPA. (Top and bottom view)

impedance matching, and peak gain, when compared to the conventional design and the original Rectangular MSPA.

The antenna characteristics for three different designs are shown in Table 1. These are the standard design, Initial MSPA, and the proposed MSPA. This side-by-side comparison shows how antenna technology has changed over time, pointing out the small variations and improvements that have been made with each new version.

**Table 1** Antenna Parameters of the Traditional Design with Initial MSPA and Novel Slot Integrated Tapered MSPA

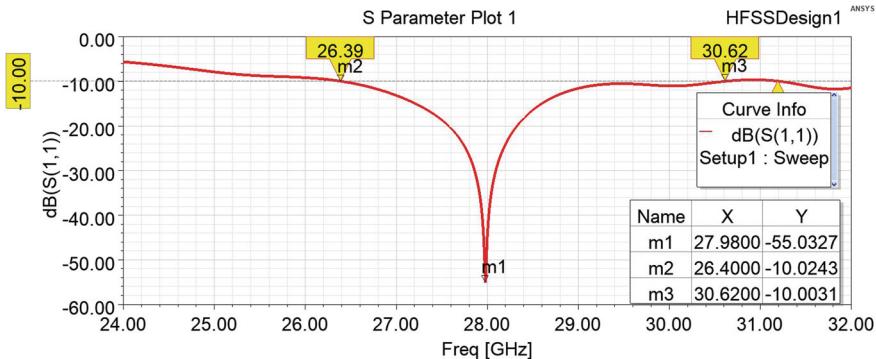
Description	Traditional antenna	Initial MSPA	Proposed MSPA
Frequency	28 GHz	28 GHz	28 GHz
Substrate	FR-4	FR-4	FR-4
Substrate thickness (h)	1.6 mm	1.6 mm	1.6 mm
Substrate width (Ws)	20 mm	20 mm	20 mm
Substrate length (Ls)	22.5 mm	22.5 mm	22.5 mm
Width of the radiating Patch element (Wp)	10 mm	12 mm	12 mm
Length of the radiating Patch element (Lp)	10 mm	11 mm	14.5 mm
Width of the partial Ground (Wg)	20 mm	20 mm	20 mm
Length of the partial Ground (Lg)	8 mm	8 mm	8 mm
Inset width (Wi)	–	2.4 mm	4.4 mm
Inset length (Li)	–	3 mm	4.4 mm
Feedline width (Wf)	2 mm	2.1 mm	2.3 mm
Feedline length (Lf)	7.5 mm	10.5 mm	10.5 mm
Length of the patch before taper (Lt1)	–	–	6.5 mm
Length of the patch after taper (Lt1)	–	–	8.94 mm
Width at the top (Wt)	–	–	4 mm
Length of the slot (Lsl)	–	–	2 mm
Width of the slot (Wsl)	–	–	3.8 mm

## 4 Parametric Analysis

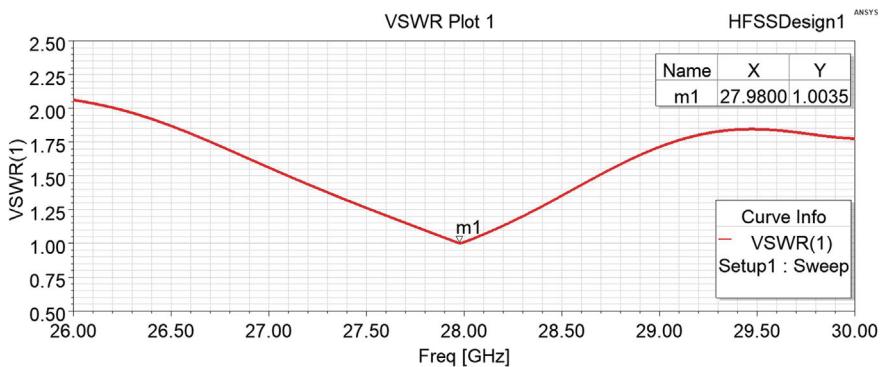
### 4.1 Initial Rectangular MSPA

The optimization capability of ANSYS' HFSS was important in enhancing the design of the conventional antenna, leading to improved performance in initial rectangular MSPA by including inset feeding and partial ground plane approaches. Simulations at the specific frequency of 27.98 GHz showed impressive performance results.

There are a lot of incredible characteristics that can be found in this system. Figure 3, demonstrates that it has a R.L of  $-55.0327$  dB along with the wide bandwidth of 4.22 GHz, while Fig. 4, demonstrates that it has a VSWR of 1.0035, and Fig. 5, demonstrates that it has a gain of 4.8 dB. And Fig. 6, illustrates the impedance matching value, which was determined to be  $49.8384 \Omega$  based on the comprehensive study conducted. This form of optimization not only makes power transmission more



**Fig. 3** –55.0327 dB of return loss and 4.22 GHz of bandwidth in the initial MSPA



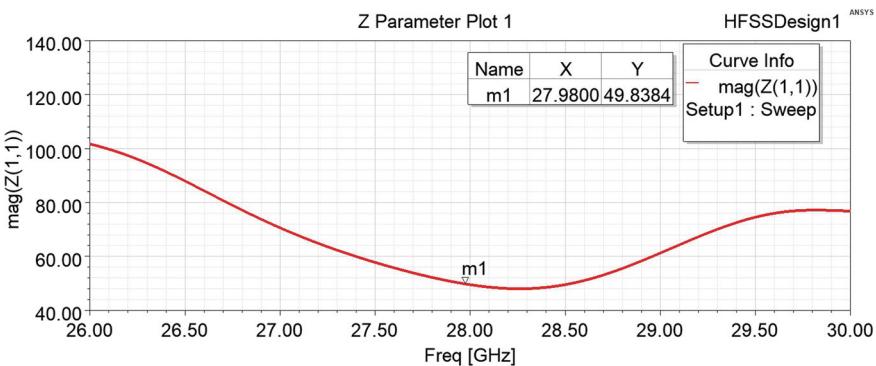
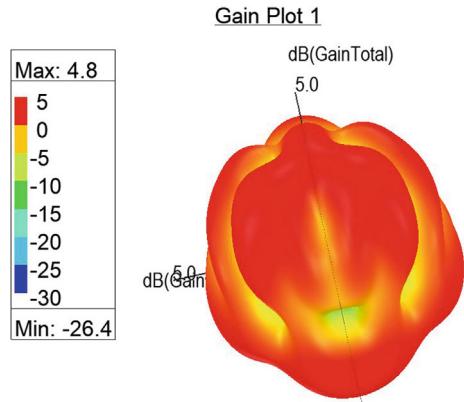
**Fig. 4** 1.0035 of VSWR in the initial MSPA

efficient, but it also lessens the number of signal reflections that are present, while at the same time increasing the bandwidth that can be used for the applications.

## 4.2 Optimization of the Slot Dimensions

Enhanced antenna design was achieved with the incorporation of tapered corners and a top slot into the patch, both of which contributed to the overall improvement. Furthermore, the optimetrics tool that is accessible inside ANSYS was employed in order to further enhance the parameters of the inset feed width, length, and feedline width. This was done in order to get the desired results. After that, the dimensions of the slot are analyzed, and it is found that the length of the slot may range from 1.5 to 2.5 mm, and the width can range from 3 to 4 mm. This information is established by the evaluation of the slot's measurements. For the purpose of the analysis, this

**Fig. 5** 4.8 dB of Peak Gain in the Initial MSPA



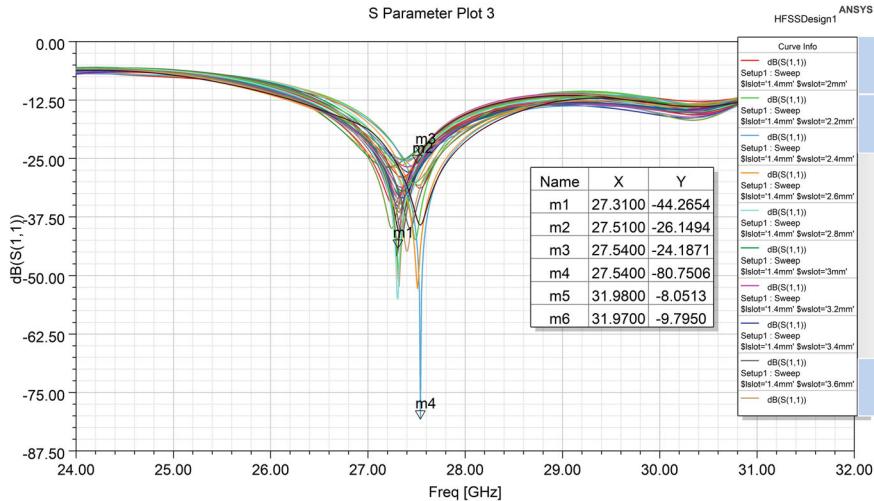
**Fig. 6** 49.8384  $\Omega$  of impedance matching in the initial MSPA

iterations used a step size of 0.1 mm. specifically, the findings of the S11 parameter and VSWR are displayed in Figs. 7 and 8, respectively.

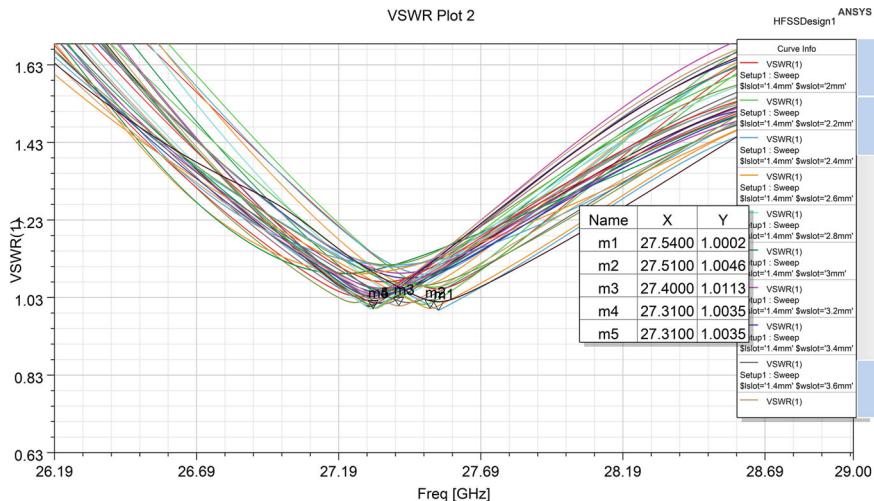
### 4.3 Optimized Novel Slot Integrated Tapered MSPA

A patch length ( $L_p$ ) of 14.5 mm, an inset gap width ( $W_i$ ) of 4.4 mm, an inset gap depth ( $L_i$ ) of 4.4 mm, and a feedline width ( $W_f$ ) of 2.3 mm were the parameters that were decided finally based on the parametric optimization. In addition, the patch gradually narrowed on both sides, starting at  $L_{t1} = 6.5$  mm and continuing to  $L_{t2} = 8.94$  mm. The measured width ( $W_t$ ) at the top of the patch was 4 mm. A slot was carefully positioned 2 mm away from the top corner, measuring 2 mm in length and 3.8 mm in breadth.

These modifications signify a notable progression in antenna design, leading to enhanced performance measures in comparison to the conventional architecture. The



**Fig. 7** Return loss obtained from optimetrics by varying slot length and width



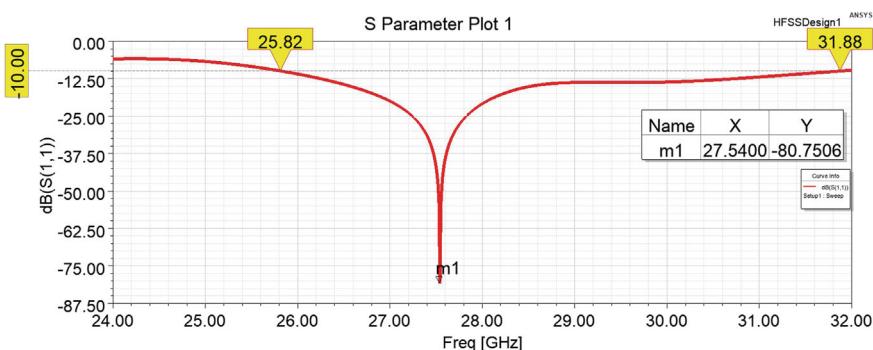
**Fig. 8** VSWR obtained from optimetrics by varying slot length and width

suggested antenna design provides improved efficiency and functionality by careful optimization and refinement of important parameters. This makes it suitable for many communication applications.

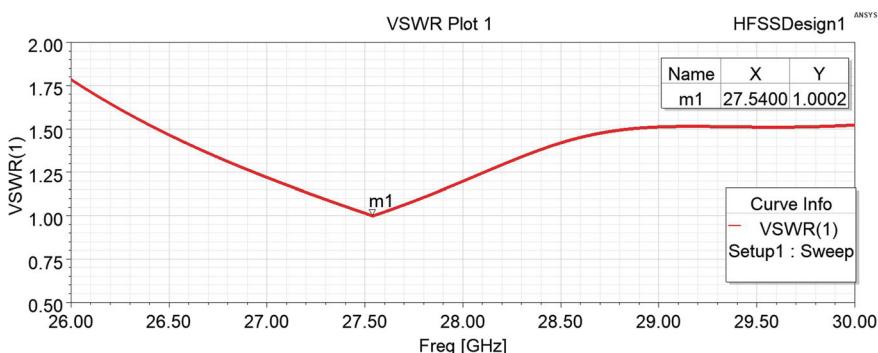
## 5 Results and Discussion

The research achieved a remarkable R.L of  $-80.75$  dB at the target frequency of  $27.54$  GHz, together with a wide bandwidth of  $6.06$  GHz, as shown in Fig. 9. The efficiency of the design optimizations and simulation approaches that were utilized in this study is demonstrated by this accomplishment. According to the return loss value that was obtained, there is a high impedance match between the antenna and its surrounding environment. This suggests that the antenna is able to transmit power effectively and that there are few signal reflections. A low return loss value implies that the antenna is able to radiate energy effectively without suffering a significant amount of loss, which ensures that it will function exceptionally well in communication scenarios that occur in the real world.

According to the findings of the research, VSWR is  $1.0002$  at the resonant frequency of  $27.54$  GHz, as can be seen in Fig. 10. When the VSWR value is close to unity, it shows that the antenna and its transmission line have an excellent impedance match. This results in minimum signal reflection and ideal power transfer efficiency.



**Fig. 9**  $-80.75$  dB of return loss and  $5.26$  GHz of bandwidth in the proposed MSPA



**Fig. 10**  $1.0002$  Of VSWR in the proposed MSPA

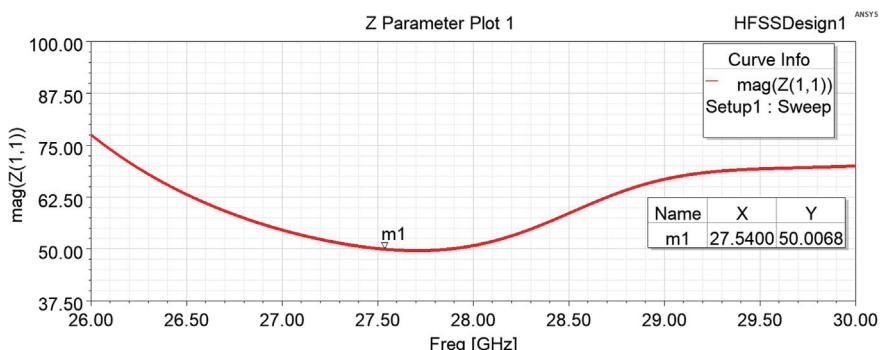
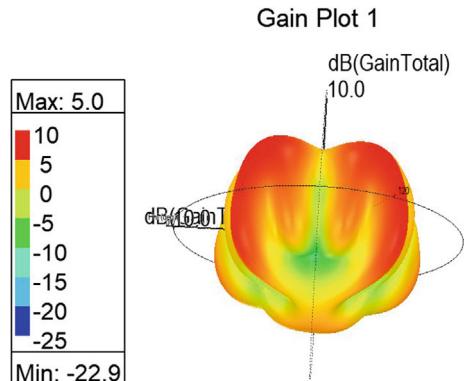
When compared to an isotropic radiator, the peak gain shown in Fig. 11 is 5.0 dB, which indicates a large increase in the amount of signal power. As a result, this reveals that the antenna is capable of successfully sending and receiving signals with increased efficiency.

In addition to that, when the resonant frequency is 27.54 GHz, the value of the impedance is equivalent to 50.0068  $\Omega$ , as seen in Fig. 12. This alignment is very consistent with the generally used standard impedance value of 50  $\Omega$  in communication systems. The strong correlation between the measured impedance and the standard impedance specification suggests that the antenna system has achieved a very efficient impedance matching.

The radiation pattern of the proposed MSPA determines its coverage area, gain, and interference rejection capabilities as shown below in Fig. 13.

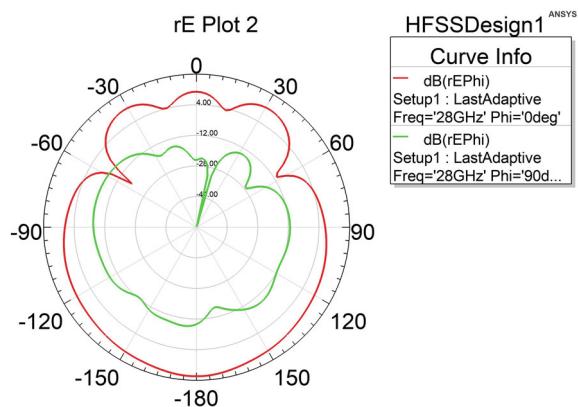
The data shown in Table 2 includes the simulation results obtained from both Initial MSPA and Proposed MSPA. The present results are juxtaposed with data from previous research papers to provide a basis for comparative analysis. The Proposed MSPA, the suggested design, has exceptional performance in terms of R.L, VSWR,

**Fig. 11** 5.0 dB of max gain in the proposed MSPA



**Fig. 12** 50.0068  $\Omega$  of Impedance Matching in the proposed MSPA

**Fig. 13** Radiation pattern of the proposed MSPA



**Table 2** Test results of the traditional MSPA versus proposed MSPA

Antenna references	Return loss (dB)	BW (GHz)	VSWR	Gain (dB)
[1]	-19.5	1.318	—	7.57
[7]	-42.134	4.841	1.0158	8.104
[8]	-22.5	5.57	1.162	5.06
[9]	-39.37	2.478	1.021	6.37
[12]	-33.37	3.56	1.04	10
[14]	-20.95	1.06	—	7.5
[15]	-21	5	—	—
Initial MSPA	-55.032	4.22	1.0035	4.8
<b>Proposed MSPA</b>	<b>-80.75</b>	<b>6.06</b>	<b>1.0002</b>	<b>5.0</b>

and bandwidth when compared to previous experiments. Since the proposed MSPA's gain is lower than that of the existing models, it will be necessary to enhance it in the future.

## 6 Conclusion and Future Work

The achieved outcome showcases the effective implementation of the antenna's capabilities across several important factors, positioning it as a very favorable option for incorporation into high-frequency communication systems. The design method is validated by the outstanding results achieved, which include exceptional return loss, near-unity VSWR, high gain, and close impedance matching to the conventional  $50\ \Omega$ . The small size, broad bandwidth, and durable performance features of this antenna make it highly suitable for many applications in radar, satellite communication, and

wireless networking at 28 GHz and in the rapidly expanding 5G communication network industry and beyond.

In the future, the gain characteristics and radiation pattern has to be improved and fabricating this MSPA design might bring about a new period of effective and dependable wireless communication at millimeter-wave frequencies, satisfying the changing needs of current communication systems.

## References

1. Hussein OD, Hussein AA (2023) Design of microstrip patch antenna for radar and 5G applications. *ESP-JETA* 3(1):74–86. ISSN: 2583-2646. <https://doi.org/10.56472/25832646/JETA-V3I2P102>
2. Kaur P, Malhotra S, Sharma M (2023) Tilted pentagon with rectangular slotted patch two-port MIMO antenna for 28 GHz 5G mm-wave band applications. In: 2023 first ICMAC, pp 1–6. <https://doi.org/10.1109/MAC58191.2023.10177095>
3. Hakim ML, Uddin MJ, Hoque J (2020) 28/38 GHz dual-band MPA with DGS and stub-slot configurations and its  $2 \times 2$  MIMO antenna design for 5G wireless communication. In: 2020 IEEE region 10 symposium (TENSYMP), pp 56–59. <https://doi.org/10.1109/TENSYMP50017.2020.9230601>
4. Nahas M (2022) A super high gain l-slotted microstrip patch antenna for 5G mobile systems operating at 26 and 28 GHz. *Eng, Technol Appl Sci Res.* <https://doi.org/10.48084/etasr.4657>
5. Soti S, Chakravarti P (2022) A compact patch antenna with DGS for 28 GHz 5G millimetre band applications. In: 2022 IEEE 7th I2CT, pp 1–4. <https://doi.org/10.1109/i2ct54291.2022.9824559>
6. Azizi MK, Rabaani K (2022) Design of microstrip antenna for 5G applications at 28 GHz. In: 2022 MMS, pp 1–3. <https://doi.org/10.1109/mms55062.2022.9825539>
7. Paul LC, Saha HK (2021) A wideband microstrip line feed slotted patch antenna for 28 GHz 5G applications, at ICECIT, Khulna, Bangladesh, pp 1–4. <https://doi.org/10.1109/ICECIT54077.2021.9641230>
8. Przesmycki R, Bugaj M, Nowosielski L (2020) Broadband microstrip antenna for 5G wireless systems operating at 28 GHz. *Electronics* 10(1):1–19
9. Kaeib AF, Shebani N, Zarek AR (2019) Design and analysis of a slotted microstrip antenna for 5G communication networks at 28 GHz. In: Proceedings of the 19th ICSTACCE, pp 648–653
10. Tirupati L, Gopalan A (2023) Simulation and comparison of novel triangle slot patch antenna with spiral loaded triangle antenna to enhance R.L and VSWR for L band applications. In: 2023 eighth ICONSTEM. IEEE, pp 1–6
11. Sanju Priya BR, Gopalan A (2023) Simulation and comparison of RF performance of the novel inverted F SIW antenna with triangular SR SIW antenna for X band applications. In: 2023 eighth ICONSTEM. IEEE, pp 1–6
12. Colaco J, Lohani R (2020) Design and implementation of microstrip patch antenna for 5G applications. In: Proceedings of the 5th ICCES, pp 682–685
13. Ganesh Reddy C, Gopalan A (2023) Simulation and comparison of square SRR with triangular slot and square SRR without slot to enhance the R.L and bandwidth performance for ITU band applications. In: Eighth ICONSTEM. IEEE, pp 1–6
14. Didi S-E, Halkhams I, Fattah M, Balboul Y, Mazer S, El Bekkali M (2022) Design of a microstrip antenna patch with a rectangular slot for 5G applications operating at 28 GHz. *TELKOMNIKA*. ISSN: 1693-6930, e-ISSN: 2302-9293 Universitas Ahmad Dahlan. <https://doi.org/10.12928/telkomnika.v20i3.231>
15. Jain R, Thakery VV, Singhal PK (2023) Employing ML models to predict R.L precisely in 5G antenna. *PIER M*, vol 118, pp 151–161

# A Review on Wireless Power Transfer Systems



Soubam Chitra Devi, Ninghoujam Juleina, Mansam Wajira,  
and Sorokhaibam Nilakanta Meitei

**Abstract** Wireless power transfer (WPT) systems are becoming popular in this modern world. This process transmits power from the transmitter coil to the receiver coil without requiring a wire connection. The transmitter coil is driven with the same resonance frequency as that of the receiving coil. Moreover, the WPT system is becoming famous in the industry all over the globe due to its increase in demand among both consumers and businesses. The main advantage of this system is that it has less effect on global warming because of its lower carbon monoxide emissions. Therefore, this paper presents a review of wireless power transfer systems. This paper also discusses wireless power transfer system techniques and their applications. In addition, challenges and future prospects for WPT systems are also presented in this paper. Finally, this paper presents a comprehensive and current bibliography of WPT systems, providing valuable resources for researchers in this field.

**Keywords** Near-field wireless power transfer · Far-field wireless power transfer · Inductive coil · Magnetic resonance-coupling · Microwave wireless power transfer · Laser wireless power transfer

## 1 Introduction

Over the last several years, there have been notable advancements in the manufacturing of portable electronic and power electronic devices for diverse applications [1]. Their defining characteristics include smaller dimensions, greater energy efficiency, and higher converted power values. There is an issue with recharging the supply battery, even if the aforementioned portable gadgets have advanced technologically [2]. These days, electronic devices like laptops, tablets, and cell phones can function for up to several dozen hours without requiring a recharge. However, these portable gadgets require continuous recharging or the acquisition of additional

---

S. C. Devi · N. Juleina · M. Wajira · S. N. Meitei (✉)

Department of Electrical Engineering, Manipur Institute of Technology, Manipur University, Canchipur, Imphal 795003, India  
e-mail: [soroniameitei@gmail.com](mailto:soroniameitei@gmail.com)

batteries. This results in higher expenses and goes against the world's environmental protection strategy. One potential solution to address this issue is the implementation of WPT technology [2–5].

A Siberian scientist named Nichola Tesla first introduced the concept of WPT at the beginning of the twentieth century [6]. In his experiment, he modelled a resonant circuit that could transfer high-frequency current into another similar type of resonant circuit. He was able to light a bulb without connecting any conducting wires to it [6].

In WPT systems, electromagnetic induction and electromagnetic radiation are the two most commonly used approaches [7]. The first approach is the most commonly used in WPT systems. It utilizes two coils. There are two coils involved: one for the transmitter and one for the receiver. Electrical energy travels from a transmitter coil to a receiver coil through a magnetic field. As stated by Faraday's law of induction [7]:

$$\varepsilon = -N \left( \frac{d\Phi}{dt} \right) \quad (1)$$

where  $\varepsilon$ ,  $\Phi$ ,  $dt$  and  $N$  represent the electromagnetic force, the magnetic flux, the change in time (in second), and the number of turns respectively. The magneto motive force (MMF) relies on  $N$  turns and the amount of current ( $I$ ) flow. Therefore, the formula for MMF is [7]:

$$mmf = N * I \quad (2)$$

where,  $mmf$ ,  $N$ , and  $I$  represent the magneto motive force, no. of  $N$  turns and current.

Due to the rapid usage of electronic devices, it even makes it complicated for a room to access them all at once, which would make a mess of the room and also expose danger with those power transferring conducting wires, which could cause fire accidents and even lead to fatal deaths [7]. Using the WPT could enhance more than all of those gadgets at the same time, without the complexity we faced with those conducting wires. Additionally, this type of power transfer offers enhanced protection. As a result, it is beneficial for the new technological world to increase its use of this power transfer [7]. The advantage of WPT is that it is simple in design, lower in cost, and can be used for short-distance applications at a lower frequency. Even though the WPT system is not a new concept, it gains popularity because of the widespread use of smartphones and other electronic devices in the twenty-first century [8, 9].

The objective of this study is to provide a comprehensive overview of WPT technology. The paper will start by providing a concise summary of the basic principles of WPT systems in order to enhance comprehension of previous breakthroughs in the area and current research. This study further demonstrates WPT system techniques and deliberates on their merits and drawbacks. Furthermore, this research delves into the challenges and future prospects facing WPT systems.

The remaining parts of the paper are arranged as follows: Sect. 2 outlines the techniques for wireless power transfer systems. Section 3 reviews the literature on

wireless power transfer systems. Section 4 describes the WPT system's applications. Section 5 discusses the challenges and future prospects. The conclusion is presented in Sect. 6.

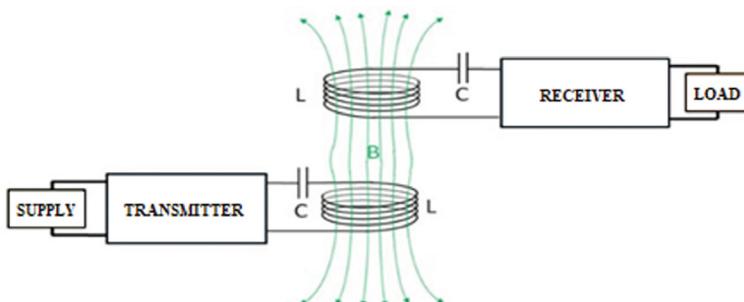
## 2 Techniques of Wireless Power Transfer Systems

There are two types of techniques for transferring power wirelessly.

### 2.1 Near-Field Technique

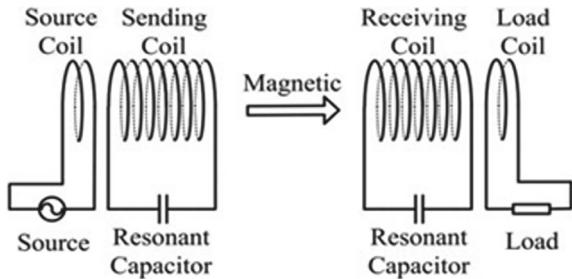
It is a short-distance wireless power transfer technique based on a magnetic field that uses inductive or magnetic resonance coupling between wire coils [10].

**Inductive Coupling WPT system:** The inductive coupling method is the most commonly used near field technique for wireless energy transfer between two coils [10]. This type of WPT is done based on the principle of mutual induction. In mutual inductance, two coils are placed near one another, and then the magnetic field in one of the coils tends to link with the other coil, allowing to induce a voltage in the second coil [11]. For example, we can see in the working concept of a transformer that primary & secondary circuits are linked by a shared magnetic flux. The transformer's primary coil side and secondary coil side remain isolated, yet they maintain a magnetic coupling. This method is a reliable and simple way to transmit power wirelessly. But the separation distance between two coils is a few mm, and the transmitted power is only a few watts. The efficiency of this method is very low. Some applications of this technology are charger pads for phones or laptops, electric toothbrushes, and wireless charging in biomedical devices like artificial cardiac pacemakers [11]. Figure 1 depicts the inductive coupling WPT system.



**Fig. 1** Inductive coupling WPT system

**Fig. 2** Magnetic resonance-coupling WPT system



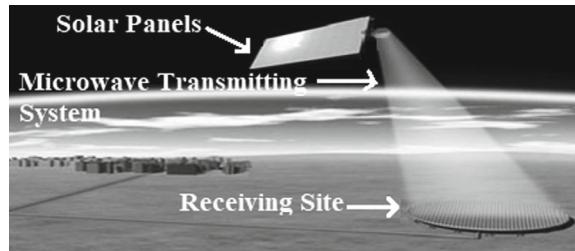
**Magnetic resonance-coupling WPT system:** Magnetic resonance coupling is one of the near-field techniques used to compensate for problems faced by non-resonance inductive coupling wireless power transmission systems. This method, which wirelessly transfers power between two coils separated by a few meters, can increase efficiency [11]. This type of coupling method applies the concept of resonance. Here, the power is transferred by magnetic flux between the two resonant circuits, one at the transmitter and another at the receiver. The two resonant circuits have a coil each. The transmitter and receiver coils are tuned in such a way that both coils have the same resonant frequency. This results in an increase in coupling efficiency. Magnetic resonance coupling is used for building mid-range wireless power transfer systems. It can be specified by distance up to 10 times the transmitting coil diameter. This type of technique is taken to be the perfect choice for home devices and appliances, as power transfer is not that much affected by humans or any other everyday objects like wood, metals, or electronic devices [10–12]. Figure 2 depicts the magnetic resonance-coupling WPT system.

## 2.2 Far-Field Technique

It is a longer-distance wireless power transfer method that uses beams of electromagnetic radiation, such as microwaves or lasers [10].

**Microwave WPT system:** Microwave wireless power transfer is one of the most efficient far-field techniques, with a separation distance range of up to 100 km and power transfer of up to 100 MW. This type of technique uses microwave frequencies ranging from 1 to 1000 GHz produced by a microwave generator [13]. The microwave energy can be converted back to electricity by using a rectenna. The rectenna consists of a dipole antenna and an RF diode connected between dipoles. The diode receives the signal from the microwaves, rectifies it, and then delivers the rectified current to smooth and control the output DC power, which it then transmits to the connected load [11]. Rectenna conversion efficiencies have been realized to exceed 95%. Solar Power Satellites (SPS) are based on this microwave wireless power transmission system. SPS converts the sun's energy into electricity in space and transmits it to

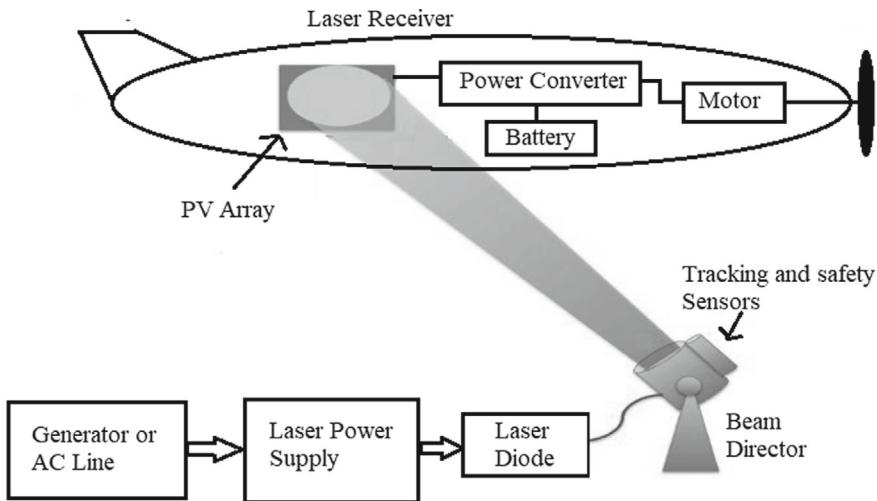
**Fig. 3** Microwave WPT system



Earth using microwaves [14, 15]. One of the major disadvantages of this method is that microwave radiation can affect humans or animals' health, leading to unwanted harmful effects on human or animal tissues if the radiation scatters from their target receiver point [11]. Figure 3 depicts the microwave WPT system.

**Laser WPT system:** A laser WPT system is also one of the types of far-field techniques that use laser beams to transmit power. In this method of power transmission, first electricity is converted into a laser beam, and then, at the receiving side, it is converted back into electricity by using photovoltaic cells [16]. On the receiver side, photovoltaic laser power converters that are specially optimized for monochromatic light conversion are applied [10]. Its power transfer is up to 100 MW and is generally used for military weapon development and space application research. The overall efficiency of laser power transfer is about 50%, which is quite low as compared to other WPT systems. However, this type of WPT is not used for electrical power transmission because of several reasons. Laser beams can easily harm humans or animals if they make it through the beam path [17, 18]. Figure 4 depicts the laser WPT system.

Generally, WPT research necessitates the exploration of multiple databases to collect relevant material. Some often-used databases for this purpose include: IEEE Xplore, ScienceDirect, Springer Nature, PubMed, Google Scholar, Engineering Village (Compendex), and Scopus. Wireless power transfer keywords and phrases were used to find relevant articles. The search approach included these keywords: WPT, inductive, resonant, electromagnetic, and magnetic power transfer Power electronics, near-field, and far-field WPT. The following inclusion criteria were used to assure article relevancy in addition to keywords: The latest articles cover WPT technology advances. English articles for understanding and analysis. WPT research publications encompass theoretical, experimental, and review articles. Publications in peer-reviewed journals or conference proceedings assure research quality and validity. The review eliminated articles that fulfilled any of the following criteria: Non-WPT articles. WPT-related articles lack empirical support or theoretical frameworks. Blogs, news, and promotional items have no peer review. Non-English articles, since translation resources were unavailable for study.



**Fig. 4** Laser WPT system

### 3 Literature Review

The WPT system is not a new idea; it has been experimented with since the end of the nineteenth century. This system's concept dates back to 1888, when Heinrich Hertz published a work. First, he succeeded in an experiment using his spark-gap radio transmitter to show evidence of radio waves. This marks the beginning of WPT towards the end of the nineteenth century [19]. Later, Nikola Tesla carried out the first famous WPT experiment in the late 1890s. A spark-excited radio frequency resonant transformer, now known as Tesla coils, was used to perform his experiment with transmitting power by means of inductive and capacitive coupling. Tesla constructed a massive coil, which he connected to a 200-foot-high mast featuring a 3-foot-diameter ball at its top. The coil receives around 300 KW of power and resonates at 150 kHz [20]. However, Tesla was unsuccessful in making a commercial product out of his experiment. However, resonant inductive coupling is now a widely used method for wireless mid-range power transfer. Resonant inductive coupling successfully installed medical devices such as pacemakers and artificial hearts in the early 1960s. Currently, it is still being used regularly in medical devices [21].

Kim et al. [22] proposed an optimal design of a WPT with multiple asymmetric self-resonators for an LED TV (150 Watt). This work's analysis employs mutual inductance calculation, and the optimal design procedure reveals the optimal placement of the self-resonators and the appropriate impedance positions. About 80% of power transfer efficiency is achieved, and a wireless power system (150 W), and a 47-in. LED TV are used to process the optimal design of this paper at 250 kHz. Shukla [23] presented the design and construction of a reliable wireless power transfer system for an embedded device. The main focus of this work is to transfer energy over a

longer distance, up to 50 cm. Here, WPT from an external source to embedded small devices located at different distances from the source is demonstrated. It reveals that the proposed system is suitable for applications like mobile phones, electric cars, unmanned aerial vehicles, robots, etc. Chaturvedi et al. [24] presented a model and analysis of high power, high-frequency wireless power transmission. Analysis is done using two models separately for power transmission up to 100 MW. The first is inductive coupling-based high power. It is estimated that the efficiency is about 70%. The other is microwave based high frequency, whose efficiency of 91.11% has been achieved. In January 2023, a space solar power prototype called MAPLE (Microwave Array for Power-Transfer Low-Orbit Experiment) was launched into orbit. It was developed by the Caltech Space Solar Power Project (SSPP), led by Ali Haji Miri. The project's aim is to harvest solar power in space and transmit the solar energy to the earth's surface. Now it is operational and has demonstrated its ability on March 3, 2023, to transmit power in space wirelessly and to beam the power to Earth for the first time [25].

Cruciani et al. [26] proposed an active shielding design for WPT systems. It is done based on the near-field coupling technique. Two active shielding designs are introduced and tested to shield the source or victim. The result shows that shielding is highly effective without affecting WPT performance. Fareq et al. [27] design a WPT based on solar energy. In this work, an inductive coupling serves as an antenna for wirelessly transmitting electrical power on a small scale. The results show that it can transfer energy up to 10 cm with 3.60 V and an efficiency of 0–10 cm (98.87–40%).

Noh et al. [28] introduced a transmitter coil system for WPT using magnetic induction, manufactured specifically for this purpose and without the use of a ferrite material. The empirical findings demonstrate that the suggested system outperforms a traditional transmitter coil in terms of its utility value. Moreover, it is found that the proposed system is costly compared to the traditional system. Mahmood et al. [29] develop a prototype energy harvesting method employing magnetic resonator coupling to operate a heart rate monitor without batteries. The suggested topologies outperformed prior research in transfer power, efficiency, and distance. Campi et al. [30] developed a magnetic resonant coupling-based WPT system to power an endoscopic capsule for direct feeding. These findings are encouraging, as the suggested WPT configuration can transmit at least 250 mW in a capsule that crosses the complete gastrointestinal track. Tanaka et al. [31] use ray-trace-based Monte Carlo simulation to demonstrate the benefits of distributed microwave power transmission (DMPT). They employed GNU Radio to construct the system. Results from experiments suggest the proposed system works. Wang et al. [32] proposed a new microwave WPT system that utilizes a frequency reconfigurable microstrip patch antenna array as opposed to a magnetic resonant coupling WPT system. This is the first time such a system has been offered. When simulated results are compared to experimental data, they show that the microwave WPT system's power transfer efficiency stays over 20% at all frequencies. The proposed method's main disadvantage is that it is more expensive than other methods. Tiefeng et al. [33] used optical software and light tools to create a laser wireless power transfer model using a power sphere. This model is used to examine the changes in energy distribution

**Table 1** Review of wireless power transfer systems

Authors and year	Methods	Important findings
Kim et al. [22], 2012	Resonant inductive coupling	It obtained that about 80% of power transfer efficiency is achieved by the proposed system
Ashukla [23], 2022	Inductive coupling	The proposed work is mainly focused on transferring energy to a longer distance of up to 50 cm. It obtained that the proposed system is suitable for applications like mobile phones, electric cars etc.
Chaturvedi et al. [24], 2017	RF inductive coupling and microwave	Proposed a high-power, high-frequency wireless power transmission. It reveals that the microwave-based high frequency and efficiency of 91.11% have been achieved
Noh et al. [28], 2016	Magnetic induction	It found that the proposed system outperforms a traditional transmitter coil in terms of its utility value
Mahmood et al. [29], 2020	Magnetic resonator coupling	The suggested topologies outperformed prior research in transfer power, efficiency, and distance
Campi et. al [30], 2023	Magnetic resonant coupling	The proposed WPT configuration can transmit at least 250 mW in a capsule that crosses the complete gastrointestinal track
Tanaka et al. [31], 2022	Microwave	It reveals that the proposed distributed microwave power transmission system has better performance than the conventional WPT system
Wang et al. [32], 2021	Microwave	It reveals that the power transfer efficiency of the proposed microwave WPT system remains consistently over 20% across various frequencies
Tiefeng et al. [33], 2023	Laser	The proposed laser-based WPT system is used to examine the changes in energy distribution inside the laser as it traverses different components

inside the laser as it traverses different components and to investigate the associated elements that have an impact. The experimental findings were very consistent with the simulated results. A review of wireless power transfer systems is presented in Table 1.

## 4 Applications of WPT Systems

### 4.1 Biomedical Devices

An inductively powered wireless pacing device is developed to realize remote cardiac stimulation. As it is wireless, there will be no health risks or patient inconveniences because of the battery life. WPT systems typically serve as the foundation for the

development of other emerging implant devices, such as implantable optogenetic devices and implantable impedance plethysmography (IGP) [34].

## 4.2 Space Solar Power Satellite (SPS)

Solar power satellites represent the largest application of WPT through microwave technology. The design envisions it as an electric power plant that orbits in the Geostationary Earth Orbit (GEO), ensuring it receives light 99% of the year. Recently, MAPLE (Microwave Array for Power-Transfer Low-Orbit Experiment) launched a space solar power prototype into orbit to wirelessly transmit power in space [15, 25].

## 4.3 WiTricity

WiTricity is a newly developed commercial technology at MIT (Massachusetts Institute of Technology) that utilizes coupled resonant objects to transmit power through the air. Devices such as laptops, DVD players, phones, and others rely on it for power. WiTricity can be installed in the ceiling, and the electronic product must have a WiTricity device to receive power [35].

## 4.4 Electric Vehicles and Mobile Phone Charging

Electric vehicles are now very popular in India and have many advantages, like zero emissions, low costs, etc. On mobile devices, wireless charging uses inductive coupling techniques. Smartphones require precise placement on a pad to initiate the charging process. The Nokia 920 was the first commercially available smartphone to have built-in wireless charging capability [1, 9].

# 5 Challenges and Future Prospects

WPT systems have promising opportunities for many applications, although they also encounter several obstacles and represent substantial future prospects. Some big problems with WPT systems are that they aren't very efficient, they have a short range, they can be affected by interference, they can cause safety issues with electromagnetic radiation, and they need to be close to each other for them to work [1, 2, 9, 15]. The primary challenges and prospects of the future are discussed below.

Challenges:

- (a) It has a high initial cost to implement.
- (b) The electromagnetic waves it emits have a biological impact on humans and animals. However, the inductive and resonant coupling methods have less impact than the laser and microwave methods.
- (c) Loss of power during long-range transmission.
- (d) The use of solar for power generation may be a problem on cloudy days because it would generate less power, which may interrupt the continuity of the power supply.

Future prospects:

- (a) Making the circuit compact using the Surface Mount Technique (SMT) for designing the primary and secondary coils of the WPT system.
- (b) Transferring power between devices, such as a mobile phone, is like sharing data.
- (c) Develop a WPT system for moving gadgets or devices such as electric vehicles, other robotic machines, or household appliances.
- (d) Implementation of WiTricity in every household.
- (e) To implement a wireless sensor to keep track of the power-generated parameter.
- (f) Implementation of solar power satellites (SPS) using the microwave power transfer method.

## 6 Conclusions

Wireless power transfer systems have been promising opportunities for many applications in this generation. From the smallest sensor to the satellite, there has been a growing desire for wireless connectivity due to its convenience and ease of maintenance. And also found that, in cases where using a wired or transmission line is complicated, wireless power transfer is the solution. Moreover, it has been determined that solar-based WPT can provide a more sustainable energy solution. Regardless of the challenges it faces, i.e., its efficiency, high cost, and distance, it is still chosen over for its merits, such as its technological advancement, integration with the IoT, and environmental awareness. Therefore, a review of wireless power transfer systems is presented in this paper and also discusses its techniques and applications, as well as the challenges and future scope of this research area. In the future, there may be research work on the development of a WPT system for moving gadgets or devices like electric vehicles and other robotic machines or household appliances.

## References

1. Detka K, Górecki K (2022) Wireless power transfer—a review. *Energies* 15(19):7236
2. Shan D, Wang H, Cao K, Zhang J (2022) Wireless power transfer system with enhanced efficiency by using frequency reconfigurable metamaterial. *Sci Rep* 12(1):331

3. Abdul-Jabbar TA, Obed AA, Abid AJ (2021) Design of an uninterrupted power supply with Li-ion battery pack: a proposal for a cost-efficient design with high protection features. *J Tech* 3(2):1–10
4. Kumar A, Singh NH, Namasudra S, Crespo RG, Moparthi NR (2024) Traffic matrix estimation using matrix-CUR decomposition. *Comput Commun*
5. Namasudra S, Lorenz P, Ghosh U (2023) The new era of computer network by using machine learning. *Mob Netw Appl* 28(2):764–766
6. Faysal MF, Islam MS (2019) Performance analysis of solar based wireless power transmission system. In: 2019 3rd international conference on electrical, computer & telecommunication engineering (ICECTE). IEEE, pp 141–144
7. Baroi S, Islam MS, Baroi S (2017) Design and simulation of different wireless power transfer circuits. In: 2017 2nd international conference on electrical & electronic engineering (ICEEE). IEEE, pp 1–4
8. Vilathgamuwa DM, Sampath JPK (2015) Wireless power transfer (WPT) for electric vehicles (EVs)—present and future trends. *Plug Electr Veh Smart Grids: Integr Tech*, 33–60
9. Sasikala G, Ranjan R (2023) Wireless power transfer for IoT applications—a review. *Self-Powered Cyber Phys Syst*, 115–139
10. Ali A, Yasin MNM, Husin MFC, Ahmad Hambali N (2019) Design and analysis of 2-coil wireless power transfer (WPT) using magnetic coupling technique. *Int J Power Electron Drive Syst* 10(2):611
11. Fotopoulos K, Flynn BW (2010) Wireless power transfer in loosely coupled links: coil misalignment model. *IEEE Trans Magn* 47(2):416–430
12. Cheah WC, Watson SA, Lennox B (2019) Limitations of wireless power transfer technologies for mobile robots. *Wirel Power Transf* 6(2):175–189
13. Tak A, Ustun TS (2016) Wireless power grid: leapfrogging in power infrastructure of developing countries. In: 2016 IEEE region 10 conference (TENCON). IEEE, pp 1274–1277
14. Jaffe P, McSpadden J (2013) Energy conversion and transmission modules for space solar power. *Proc IEEE* 101(6):1424–1437
15. Maqsood M, Nauman Nasir M (2013) Wireless electricity (Power) transmission using solar based power satellite technology. In: *Journal of physics: conference series*, vol 439, no 1. IOP Publishing, p 012046
16. Baghel AK, Behera C, Amalraj S, Singh A, Nayak SK (2022) Comparative study of wireless power transfer and its future prospective. In: *Smart and intelligent systems: proceedings of SIS 2021*. Springer Singapore, pp 219–228
17. Chhawchharia S, Sahoo SK, Balamurugan M, Sukchai S, Yanine F (2018) Investigation of wireless power transfer applications with a focus on renewable energy. *Renew Sustain Energy Rev* 91:888–902
18. Anand P, Pandiarajan R, Raju P (2015) Wireless power transmission to UAV using LASER beaming. *Int J Mech Eng Res* 5(1):1–6
19. Morsi R (2020) Analysis and design of communication systems with wireless power transfer. PhD diss., Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
20. Saleem B, Badar R, Manzoor A, Judge MA, Boudjadar J, Ul Islam S (2022) Fully adaptive recurrent Neuro-fuzzy control for power system stability enhancement in Multi Machine System. *IEEE Access* 10:36464–36476
21. Ben Amar A, Kouki AB, Cao H (2015) Power approaches for implantable medical devices. *Sensors* 15(11):28889–28914
22. Kim J, Son H-C, Kim D-H, Park Y-J (2012) Optimal design of a wireless power transfer system with multiple self-resonators for an LED TV. *IEEE Trans Consum Electron* 58(3):775–780
23. Shukla DA (2022) Design and construction of a reliable wireless power transfer system for an embedded device: with emphasis on industrial applications
24. Chaturvedi S, Tripathi ES (2017) Modelling & analysis of high-power, high-frequency wireless power transmission. *Int Res J Eng Technol (IRJET)* 4(05):2732–2737
25. Caltech (2023) Caltech's space solar power demonstrator wirelessly transmits power in space. <https://www.caltech.edu/about/news/in-a-first-caltechs-space-solar-power-demonstrator-wirelessly-transmits-power-in-space>. Accessed 01 Jun 2023

26. Cruciani S, Campi T, Maradei F, Feliziani M (2019) Active shielding design for wireless power transfer systems. *IEEE Trans Electromagn Compat* 61(6):1953–1960
27. Fareq M, Fitra M, Irwanto M, Syafruddin HS, Gomesh N, Rozailan M, Arinal M, Irwan YM, Zarinatul J (2014) Wireless power transfer by using solar energy. *TELKOMNIKA (Telecommun Comput Electron Control)* 12(3):519–524
28. Noh E, Ko KH, Kim K (2016) Transmitter coil system without ferrite in wireless power transfer. *Electron Lett* 52(5):392–393
29. Mahmood MF, Mohammed SL, Gharghan SK, Al-Naji A, Chahl J (2020) Hybrid coils-based wireless power transfer for intelligent sensors. *Sensors* 20(9):2549
30. Campi T, Cruciani S, Maradei F, Feliziani M (2023) A new transmitting coil for powering endoscopic capsules using wireless power transfer. *Electronics* 12(8):1942
31. Tanaka Y, Hamase H, Kanai K, Hasaba R, Sato H, Koyanagi Y, Ikeda T et al (2022) Simulation and implementation of distributed microwave wireless power transfer system. *IEEE Trans Microw Theory Tech* 71(1):102–111
32. Wang H, Deng L, Luo H, Du J, Zhou D, Huang S (2021) Microwave wireless power transfer system based on a frequency reconfigurable microstrip patch antenna array. *Energies* 14(2):415
33. He T, Zheng G, Wu Q, Huang H, Wan L, Xu K, Shi T, Lv Z (2023) Analysis and experiment of laser energy distribution of laser wireless power transmission based on a power sphere receiver. *Photonics* 10(7):844. MDPI
34. Zellagui M (2021) Introductory chapter: overview of wireless power transfer technologies. *Wirel Power Transfer—Recent Dev Appl New Perspect*
35. Rajper SZ, Albrecht J (2020) Prospects of electric vehicles in the developing countries: a literature review. *Sustainability* 12(5):1906

# Application of Big Data to Traffic Generated in Mechanisms Containment on Optical Burst Switching Distributed Networks



Oscar Corredor, Jannet Ortiz, Luis Ballesteros, Sergio Bermúdez, Carlos Enrique Montenegro-Marin, and Ruben Gonzalez-Crespo

**Abstract** In the telecommunications industry different transmission media are used which allow sending and receiving information with particular characteristics and structures that provide lower or higher transmission speed, distance, available bandwidth, frequency spectrum for propagation through the channel and data quality; The model proposed in this research uses fiber optic networks based on optical packet switching to improve network performance by sending bursts that are optically switched at the nodes generating high volumes of data that will allow to know the behavior of mechanisms containment in optical burst switching (OBS) distributed networks. From there, the proposed methodology considered transmission parameters that offered alternatives for massive data transport by deepening the interconnection of optical lines that contribute to improve the use of wavelengths due to their high speed in the process of establishing and releasing routes for bursts in transit; the simulation of the interconnection monitoring and its particularities were focused on measurements implemented from the mechanisms containment to evaluate the behavior of the transmission parameters and to recognize the potential of the optical fiber, which resulted in a 250% increase in transmission speed and a management to

---

O. Corredor · J. Ortiz · L. Ballesteros · S. Bermúdez  
Universidad Cooperativa de Colombia, Bogotá, Colombia  
e-mail: [oscar.corredor@campusucc.edu.co](mailto:oscar.corredor@campusucc.edu.co)

J. Ortiz  
e-mail: [jannet.ortiz@campusucc.edu.co](mailto:jannet.ortiz@campusucc.edu.co)

L. Ballesteros  
e-mail: [luis.ballesterosr@campusucc.edu.co](mailto:luis.ballesterosr@campusucc.edu.co)

S. Bermúdez  
e-mail: [sergio.bermudez@campusucc.edu.co](mailto:sergio.bermudez@campusucc.edu.co)

C. E. Montenegro-Marín  
Universidad Distrital Francisco José de Caldas, Bogotá, Colombia  
e-mail: [Cemontenegrom@udistrital.edu.co](mailto:Cemontenegrom@udistrital.edu.co)

R. Gonzalez-Crespo (✉)  
Universidad Internacional de La Rioja, Madrid, España  
e-mail: [ruben.gonzalez@unir.net](mailto:ruben.gonzalez@unir.net)

support the requirements for handling large volumes of bandwidth under the OBS technique.

**Keywords** Switching techniques · Data mining · Edge node · Fiber delay lines

## 1 Introduction

Currently, the high volume of information circulating in the networks generates exorbitant bandwidth consumption in technological platforms. Such data represents classified, secret, selected, restricted, encrypted, and social networks information that goes from one node to another through the existing interconnections in the world [1] using different transmission media used by communication systems. In this sense, one of the most used transmission media is optical fiber that transports information by sending light signals (electromagnetic waves) that travel through glass filaments, being a great alternative great alternative because it transports considerable volumes of traffic with different switching techniques such as OBS that supports encapsulation in bursts and that allow a combination of packets and circuits at high speed, but with loss problems because they cannot be stored in a node due to lack of optical memory [2, 3]. However, this technique allows flow control, admission, data reception, feedback, and traffic at the nodes [4] being the delays, wavelengths and packet counting, the object of study of several authors referenced in [5, 6] which include a large number of nodes and immeasurable information traffic. In this paper.

## 2 Materials and Methods

Optical fiber is a transmission media that provides higher speed for sending data because the light pulses used for packet traffic can be configured either through optical packet switching or optical burst switching [7]; for this reason and supported by some precedents [8–12], in this chapter will perform a systematic literature analysis in certified databases in order to review the background of the object of study.

### 2.1 *Packet Routing*

Routers allow data bursts to be routed to transfer information statically or dynamically and in this way obtain an efficient and effective containment process to reduce congestion through protocols, making the system recognize the best alternative routes and rules toward a final destination [13].

Here it is important to identify that as this type of element is implemented, a large capacity of resources is consumed, but at the same time it is an alternative that offers the use of the same network for the diversion of bursts by routing.

## 2.2 *Delay times*

In fiber optic networks, there are no data storage devices during information transmission [14] and neither are there buffers on the market capable of storing data in the form of “visible spectrum radiation”.

It is important to note that buffers are used at times of network congestion, so they have a time limit at that peak and can be configured to be fixed to the network (serial) or variable (parallel) trusting it the action of taking and delivering data through the buffer architecture until the information is successfully transmitted.

But thanks to the existence of transmission delay times, it was possible to create fiber delay lines (FDL), which simulate a static random access memory or RAM that to a certain extent delays the signal travels through the network producing defined delays in the optical fiber that compensate the delay differences.

## 2.3 *Wavelength Conversion*

With the design and construction of combinational optical terminal devices with several inputs and outputs (multiplexers), it is possible to offer stable channel capacity for sending information through the use of the same link, since there are several wavelengths that allow navigation of two or more bursts [14].

Likewise, one of the key elements regarding the containment of bursts are the wavelength converters which convert an incoming wavelength signal in the network to a different one at the output of the device, enabling the reuse of wavelengths for transporting burst between nodes.

It is important to point out that wavelength converters present an alternative for optimizing the network performance with a great advantage since they greatly reduce the loss of bursts; however, it is a great cost to obtain the best quality in data transmission since it is necessary to have a converter at each switch port.

## 2.4 *Interconnection Communication*

For the interconnection of network segments it is necessary to establish communication protocols by means of rules that allow the fast and lossless transmission and reception of data travelling on electromagnetic waves originating in the laser.

This communication is done through ARP (Address Resolution Protocol), which translates IP (Internet Protocol) addresses to MAC (Media Access Control Address) type in order to capture the time between the host, the Edge and the Core.

### 3 Results and Discussion

Optical burst switching is a concept that attempts to eliminate the conversion of optical to electrical signals and/or vice versa at each node in the network; however such conversion is used at a common switching destination header [16, 17]; This is because it organizes and groups packets containing similar characteristics to be sent in bursts from one node to another (edge node-border node-end node), making efficient use of network resources and flexible handling of such technology because it adapts to variable traffic dynamics [18], also allowing information flow over long distances [19].

#### 3.1 *Control Mechanisms Containment in Fiber Optic Networks*

In the evolution of information transmission systems, different variables are analyzed with the purpose of reducing network saturation, data loss and delay; for this reason, it is important to review the necessary mechanisms that allow addressing these variables focused on mitigating signal contamination, minimizing destabilization and incidents that may occur in the network [20]. Therefore, the control mechanisms will be studied in detail hoping to identify characteristics for the optimal operation of the network.

#### 3.2 *Delay Lines*

Normally in packet-switching networks, through the use of buffers, there is the possibility of saving information in time intervals while it is allowed to send the data that is waiting in the queue; However, thinking about storage in these optical system without the use of memories that can store light seems impossible, so other types of mechanisms for data or packets delay must be sought.

It is there where the FDL are involved and configured as device equivalent to a computer's main memory (like RAM) where the data is stored, although not with the same functionalities; In other words, a FDL is a significant delay time that normally is for a group of information [20].

An FDL can be fixed or variable and can be built using a simple piece in such a way that the delay is directly proportional to the length of the line, achieving a time

of 5  $\mu$ s per kilometer of fiber traveled. Fixed FDLs provide series-connected delays lines within a network, and variable FDLs are connected in parallel. In [20] we can see an overview of each of the delay line configurations.

Likewise, each one of the FDL can be classified according to its architecture as pre-feed, feedback and hybrid architecture; The first, keeps an output port of a switch interconnected with an input port of the next switch; the second one keeps the burst circulating in a delay line through a switch until it is transmitted or, on the contrary, it decides to discard it and the third configuration simultaneously integrates the interconnected system of ports allowing the burst to circulate while it is transmitted [20]. In [20] some FDL configurations are shown according to their classification (Pre-feed and dedicated FDL or Pre-feed and shared FDL).

### ***3.3 Deflection Routing***

Generally, in an OBS network, elements or network nodes are added that allow efficiently establishing containment parameters which respond to congestion problems; however, it is important to point out that this procedure increases the use of system resources, causing it to look for routing alternatives, going from one channel to another and travelling through the network until it reaches its destination, but in this journey the destination address could be lost and the network would take more circulation time to find it again or possibly even lose this information. This traffic technique focuses on the use of shared buffers through complete links within the network where each central node must receive the information periodically in the different network links, which could even aggravate the performance of the network and increase the arrival time at your final destination.

### ***3.4 Wavelength Conversion***

Fiber Optics provide the ability to use several channels on the same link, since it allows the use of multiple wavelengths to send information; however, when the bursts are sent, they can try to travel on the same channel but with the current capabilities of fiber optics, the burst can be transported at different wavelengths to avoid interference [21] using wavelength converters, which they allow an incoming wavelength to be converted to a different output wavelength, allowing a high percentage (40% maximum) to reuse the wavelengths that in turn can carry different packets through the fiber optic network.

Depending on their configuration, wavelength converters can be found in two archetypes: fixed and adjustable; the first allows an incoming channel to be connected to one or more output channels while maintaining the same wavelength at both the input and output. The second converts the incoming wavelength to a different

outgoing wavelength; however it has a range limit to perform the wavelength conversions; it is important to note that these converters are used in the core nodes [20], As shown in [20] there are Waveform converter at the input ports and Partial waveform converter.

Normally, this technique is not used individually since its effectiveness is low, so it is used in conjunction with another means of containment such as FDL.

### ***3.5 Burst Segmentation***

In an OBS Network, burst containment control techniques often resend the packets discarded in the transmission processes, which generates reprocessing, high use of resources and loss of time, congesting the communication of the nodes in the entire network. To solve this problem, a burst segmentation is used overlapping other segments of different bursts, pretending not to eliminate packets in their entirety; In addition, it provides head-dropp or tail-dropp bursts options, resulting in high quality transport. However, in segmentation control and in edge nodes, segmentation recovery is complex, as shown in [20].

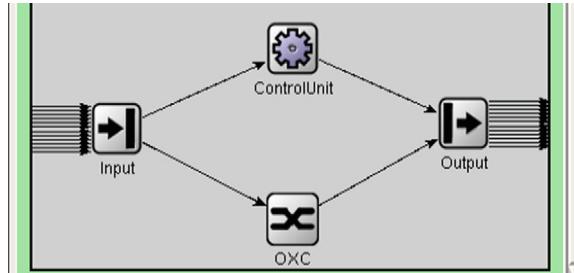
### ***3.6 Design, Transmission and Simulation Parameters***

OMNeT ++ has basic tools composed of C ++ class libraries for the creation of modules, channels and interfaces with the purpose of simulating network environments or contexts, but which by itself does not provide system architecture components or queue simulations, so it is necessary to integrate it with other tools that complement the simulation engine such as INET Framework or Mobility Framework.

In order to specify the simulation, it was necessary to follow a structure of basic steps that allowed the construction of the model to be implemented in such a way that the necessary hierarchy for communication and network connection between the modules and related sub-modules through the NED files is contemplated; As a first step, the system map was built from the communication of the modules between them, forming the composite module; the second step, the model was defined in the NED language as an operating structure using the OMNET ++ graphical editor; As a third step, the model components were activated using the resources and software libraries; In the fourth step, we proceeded to obtain the “.ini” and “config” file with the parameters for the different simulations; As a fifth step, the program was run to create the connection between the code and the user interface provided by OMNET ++ ; and finally the results with vector types values were obtained for analysis.

Now, it is important to point out some of the components of OMNET ++ to make visible the management, execution and simulation of the network, among them are the nodes (central, input and output edge), the libraries, compiler, executables and user interface, among others.

**Fig. 1** Core sub-components



### 3.7 Central Node

The central node consists of four modules which are: input module, control unit module, optical cross-connect interface (OXC) module and output module. The input module divides the optical fibers into two parts, control channels connected to the control unit and data channels connected to the OXC [11].

So, if the packet arrives at the control header, it is sent to the control unit on a dedicated channel; but if the data burst arrives, it is subsequently sent to the OXC through which it passes through the entire optical path. Additionally, the control unit module is responsible for processing the control header and forwarding it to the particular output link according to the routing table. The functionality of the control unit is implemented by means of the logic control submodule that is in charge of routing, control and updating of the header information, generating the wavelength [11] (Fig. 1).

### 3.8 Edge Node

It is divided into two types: input edge node and output edge node, which acts as an input node when traffic enters the network and as an output node when it removes the said traffic [11].

However, when the input node function is executed, the module is engaged in building the packets into bursts and scheduling the transmission of the bursts on the output channel. Packets only go through this module if their destination and class of service are valid for sending through the network as shown in edge subcomponents of [12]

### 3.9 OMNET ++ User Interface

“Tkenv” allows simulation to be performed at the highest level hierarchy during this stage, supporting visualization, debugging and providing a real-time experimental

environment during execution. “Cmdevn” is a small, fast and portable interface that can be used on platforms such as DOS and/or UNIX. “Tvenv” is an interface that can be manipulated via character, mouse or windows overlay and that allows interactive execution, tracing and debugging. This interface is recommended to be used in presentations because more detailed images of the simulation are obtained; some features of this interface are: the modules have separate windows for the text output; simulation messages can be viewed in a separate window and the simulation can be restarted instantly with detailed model data.

## 4 Architecture of an OBS Network

The OBS architecture must be able to tolerate a high demand for bandwidth, adapting to the circulating traffic for the transport of information from the grouping of data oriented to a destination under characteristics or criteria established in the processing, signaling and sending of the packets.

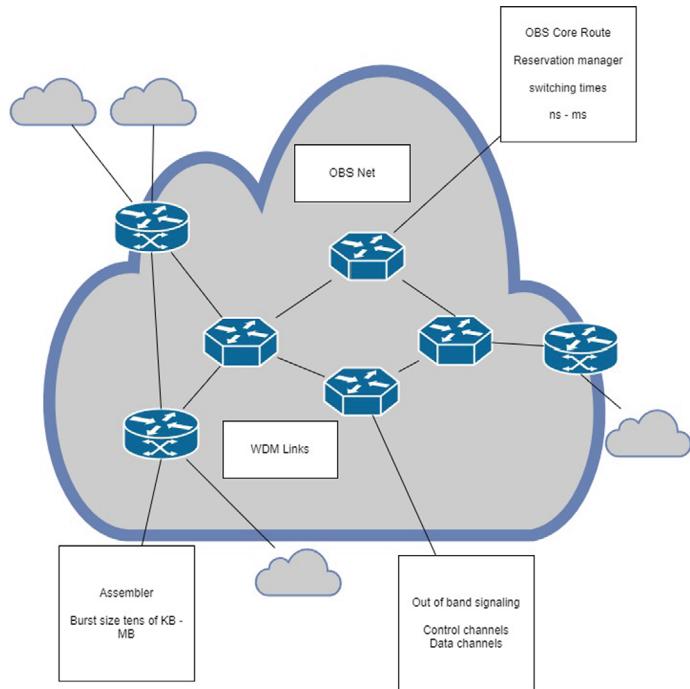
The basic architectural environment of an OBS Network is presented in Fig. 7, where it can be seen that the bursts sent are emitted from the emitter modules until an assembly phase in which the network, with sophisticated programming, executes the conversion process from electronic to optical signals, certifying that the bursts reach the border nodes to be transmitted according to the routing tables in a transparent way [10] (Fig. 2).

This architecture is designed to solve, execute and organize the procedures, operations and information formats in a functional way; however, for an OBS network, it is important to have control mechanisms that allow decongesting the nodes through classification criteria, being responsible for the routing of the bursts, transmission link and traffic flow through the network. Therefore, in this work, different models of distributed traffic containment systems were used in the OMNeT ++ simulator to give capacity to bandwidth, reduce data loss and packet transmission delays.

It is important to highlight that the simulation of an OBS network is necessary due to the demand for bandwidth in optical backbones as an alternative for fast, efficient and reliable traffic in the transport of packets, which is why it was included in the simulation of burst switching, different control mechanisms for the decongestion of nodes due to excess information.

### 4.1 Data Collection and Big Data Application

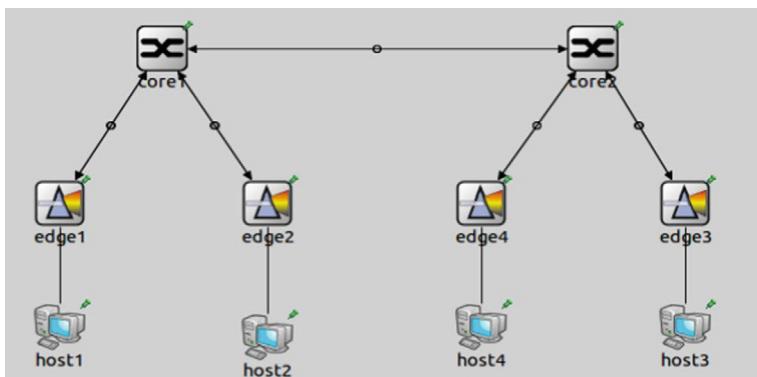
In papers such as [22] that comprises three modules, use a big data in the data processing and analysis module. In this research the OMNeT ++ software allows the creation of a network model from modules designed to be combined and communicate with each other and to simulate real-time routines and components in computer networks [5, 6]; in this research, the scheme used for an OBS Network is presented



**Fig. 2** Basic architecture of an OBS network

in Fig. 3, where the design of its physical topology has two Core, four Edge and four Hosts.

The process of applying Big Data to the data obtained from the simulation, was carried out with the free version of Power Bi Desktop software. The first step was



**Fig. 3** Physical topology of an OBS network

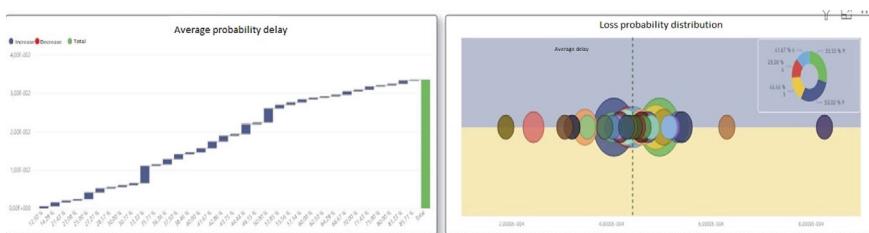
to import, from an Excel worksheet, the data obtained from the simulation of each of the containment control mechanisms; then the data resulting from the said table were transformed with the purpose of adapting them for the analysis. This procedure is carried out for each one of the columns contained in the Excel worksheet; subsequently, data processing begins where, to refine the data query, the rows containing null values are filtered, since these behave as non-existent data and possible errors for the analysis; then, the visualization process is started based on the data obtained; finally, the Power BI Desktop panels are customized for a visualization according to the data that is to be presented.

The results of the simulations of the containment control mechanisms from data processing in Power BI Desktop software are shown below, organized by delay lines, wavelength conversion, and routing deflection; additionally, a simulation without contention mechanism was carried out, which provides a line comparison in sending and receiving packets, average delay and probability distribution of data.

## 4.2 By Delay Lines

Figure 4 shows three panels: the upper figure is a set of bars showing the ratio of bursts sent vs. bursts received during set time intervals in 10 ms ranges, starting at 500 up to 1500 ms; in the figure on the lower-left you can see the average delay vs. percentage probability of loss and in figure on the lower right, the distribution trends of probability of loss.

The bar graph provides the information supplied to the network simulator regarding the packets sent (green, with an average of 10 packets) vs. packets that arrive at their destination (blue, with an average of 6 packets received) in 10 ms intervals starting at 500 up to 1500 ms and taking into account that a time is required to stabilize the system. For the simulation test, 100 packets were sent, made up of groups of 3 to 21 data packets, providing an average fiber line delay equal to 440  $\mu$ s and a packet loss probability of 44%, meaning that if for example two groups of packets arrive at a node, only one of them manages to pass successfully while the other must wait until it finds an available line to continue; however, if it takes longer than appropriate, other packets could arrive and again limit the use of output lines or



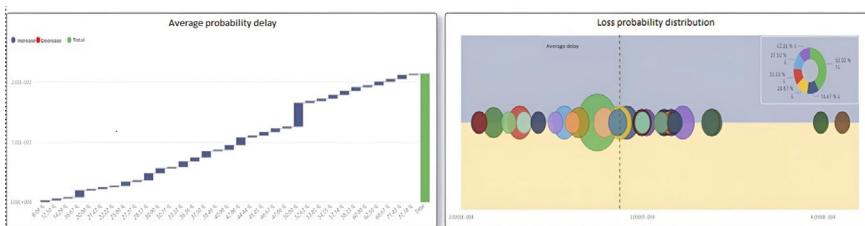
**Fig. 4** Data analysis by delay line control

they could be lost. The waterfall plot (left-bottom) provides the ratio of probability loss (x-axis) vs. delay (y-axis), revealing an ascending line from its beginnings to the entire simulation (86%), which means a cumulative impact in time in a negative sequential manner; in other words, the longer packets take to be delivered to their destination, the more likely they are to be lost. The scatter plot reveals the relationship between the variables probability of loss vs. delay variables for each of the set of processed packets. There it can be seen that a large amount (75%) is concentrated around the 440  $\mu$ s, without distinction of the number of packets within each group, which leads to ratify that the fiber line control mechanisms can be a seasonal solution over time while the technology provides optical storage devices.

### 4.3 By Wavelength Conversion

Figure 5 also presents three panels: The figure at the top shows the ratio of bursts sent vs. bursts received during time intervals established in ranges of 10 ms, starting at 500 up to 1500 ms; in the figure on the lower-left part, the average delay vs. percentage of probability of loss can be observed and in the figure on the lower-right part, the distribution trends of probability of loss.

The bar chart allows visualizing the information provided by the network simulation system regarding the packets sent (green, with an average of 10 packets) versus packets arriving at their destination (blue, with an average of 6 packets received), for periods of time in the range of 10 ms, from 500 to 1500 ms, with a network stabilization time; In this way, 100 packets made up of groups of 4 to 19 data packets were sent, with an average by control of the wavelength conversion mechanism equal to 277  $\mu$ s and a packet loss probability of 41%, resulting in the planned channels on the horizon will have an average of 277  $\mu$ s of waiting while the next wavelength is released and the system is updated to receive the next burst; however, if the expectation exceeds the time limit to continue to its destination, the loss data will be inevitable. The waterfall graph (left-bottom) provides the relationship between probability of loss (x-axis) vs. delay (y-axis), showing an ascending line from its beginning to the entire simulation (78%), observing in a first part (0 to 50%) a probability with an ascending trend and not as pronounced, but equally with a negative cumulative impact over time; in other



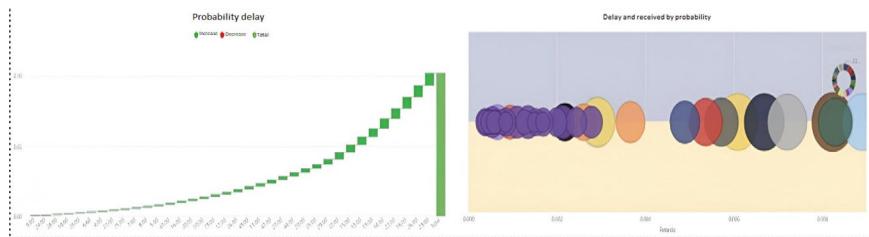
**Fig. 5** Data analysis by wavelength conversion control

words, the disposition of the packets to continue their route is more viable than the one used by the FDL control mechanism, but, with the same risk of probability of data loss. The scatter plot for the wavelength control mechanism provides the correlation, on the horizon, of the variables probability of loss vs. delay for each one of the set of processed packets, where it is observed that a large amount of data is concentrated in delay times from 200 to 277  $\mu$ s, which in this case would be 80% of the processed packets, without distinction of the number of packets within each group; This allows verifying that the use of the wavelength conversion control mechanism is a great alternative for sending information in a fiber optic network.

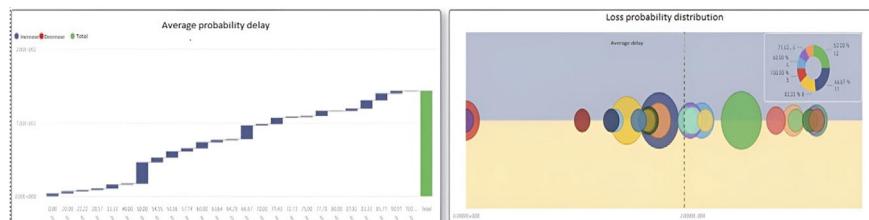
#### 4.4 By Route Deflection

Figure 6 shows three panels that allow us to observe: the bar figure provides the number of bursts sent vs. bursts received during time intervals established in ranges of 10 ms, starting at 500 up to 1500 ms, with a stability time of network; in the lower-left part, the percentage probability of loss versus average delay is presented and in the lower-right part, the distribution trends of probability of loss.

The bar chart shows the information resulting from the simulation in the OMNeT ++ software, regarding the packets sent (green, with an average of 10 packets) versus packets that arrive at their destination (blue, with an average of 7 packets received), for periods of time in a range of 10 ms, starting at 500 up to 1500 ms, with a network



**Fig. 6** Data analysis by route deflection control



**Fig. 7** Data analysis without control mechanism

stability time. In this case, 100 sets of packets were sent, grouped from 4 to 19 data bursts, with an average delay, according to the route deflection mechanism, equal to 1 ms and a packet loss probability of 24%, meaning that the packets can be affected by the delay time because they consume more time to navigate the network to reach their destination, but present less probability of loss; However, if the arrival time is exceeded, more than expected, the packets may get out of order and cause instability in the network. The waterfall plot (left-bottom) shows the relationship between probability of loss (x-axis) vs. delay (y-axis), revealing a line with a slight rise from its beginning to the totality of its data, which means a cumulative impact the time in a leisurely negative sequential manner; in other words, as packets are routed, they take longer to be delivered to their destination, but are less likely to be lost. The scatter graph reveals the relationship between the variables: probability of loss vs. delay for each of the representations of the processed packets, where it is observed that 98% (of 100 bursts sent) have an average of less than 50% probability of loss, regardless of the number of packets within each group, which leads to recognizing that the route deflection mechanism has a higher probability that the data sent will reach its destination but with a longer residence time in the network.

## 4.5 *Simulation of OBS Network Without Containment Mechanism*

In Fig. 7, three panels are displayed that allow observing in the upper part, the proportion of bursts sent vs. bursts received during time intervals set in the range of 10 ms, starting at 500 up to 1500 ms with a network stabilization time; in the lower-left part the average delay versus percentage probability of loss and in the lower-right part the distribution trends of probability of loss.

The bar chart shows the information resulting from the network simulation system regarding the packets sent (green colour, with an average of 7 packets) vs. packets that arrive at their destination (blue colour, with an average of 2 packets received), for periods of time in a range of 10 ms. For this simulation without containment mechanism, 100 sets of packets were sent, consisting of groups of 2 to 15 bursts, presenting an average delay without the control mechanism equal to 185  $\mu$ s and a packet loss probability of 65%, which possibly indicates a delay time from packet delivery to the initial node in the network path. The waterfall graph (left-bottom) shows the relationship between probability of loss (x-axis) vs. delay (y-axis), revealing a cascading ascending line from its beginning to the entire simulation (100%), observing a cumulative 50% very close to the perpendicular intersection of the graph, which means a gradual cumulative negative impact on the said relationship; In other words, the loss of bursts is very high, because in a fiber optic network there are no information storage devices while they are routed to their destination, therefore, the transmission is affected throughout its entire journey. The scatter plot reveals the relationship between the probability of loss vs. delay variables for each of the groups of packets

**Table 1** Comparative table of results

Control mechanism/variable	Delays	Packets loss	Average probability loss
FDL	440	44	25
Wavelength conversion	270	41	20
Route deflection	1 ms	24	2
without containment mechanism	185	65	73

processed, in which only 27% of the 100 bursts sent have a 50% probability of failure arrive at your destination without inconvenience.

Table 1 below compiles the information provided from the control mechanisms simulated in OMNeT ++ software and analyzed in Power BI Desktop software.

## 5 Conclusion

By evaluating the simulations performed in OMNeT ++ software in the time domain in an OBS network against the probability of data loss, it was possible to observe that distribution of the burst from the route deflection mechanism between the network nodes ensures a reliable transmission, minimizing the probability of packet loss and establishing, at the same time, the highest amount of data delivery to its destination with the disadvantage of providing the system with the longest delay time; Also, it was found that the wavelength conversion mechanism provides an average distribution reading both in delay time and in probability of loss of packets sent.

Likewise, even though OBS networks are constantly losing information segments with which the quality and performance of this network is affected, the impact of the use of control mechanisms demonstrates that it is still a high-speed network designed as an alternative traffic flow model in real scenarios. It is important to highlight the application of the use of Big Data for stochastic analysis from the simulations, since it allowed establishing the benefits of using containment control mechanisms against a network without the said procedure, as well as to discover of the behavior pattern of each of them, with correlation of variables and probability distribution of data loss according to the time of each of the control systems used.

## References

1. Asteasuain F, D'Angiolo F, Dubinsky M, Pazos F, Kwist I, Loiseau M, Calonge F (2020) Smart applications on the internet of things and big data: a rigorous approach. In XXII Work Res Comput Sci (WICC 2020, El Calafate, Santa Cruz)
2. Freire LA (2010) Behaviour Of OBS (Optical Burst Switching) Networks With TCP Traffic," Quito

3. Almeida Freire AL (2011) Behavior of OBS (Optical Burst Switching) networks with TCP traffic (Bachelor's thesis, QUITO/EPN/2011)
4. Farahmand F, Zhang Q, Jue JP (2004) A feedback-based conflict avoidance mechanism for optical burst switching networks. In: Proceedings Workshop Optical Burst Switching
5. Guevara Ortiz KJ, Díaz Erazo YR (2015) Impact of contention control mechanisms in distributed OBS networks
6. Guevara-Ortiz K, Bermúdez-Quintero S, Puentes A, Villabona-Mujica N, Rosero LB, Suarez-Rivero D, Ortiz-Aguilar J (2018) Impact of control mechanisms containment in optical burst switching distributed networks. In: 2018 IEEE Colombian Conference on Communications and Computing (COLCOM) (pp 1–6). IEEE
7. Qiao C, Yoo M (1999) Optical burst switching (OBS)-a new paradigm for an optical internet. *J High Speed Netw* 8(1):69–84
8. Coulibaly Y, Abd Latiff MS, Selamat A (2009) A novel routing optimization in optical burst switching networks. In: 2009 2nd International Conference on Communication Theory, Reliability, and Quality of Service, CTRQ 2009 (pp 108–112). <https://doi.org/10.1109/CTRQ.2009.9.10>
9. Dutta, MK, Chaubey, VK (2012) Comparative analysis of wavelength conversion and segmentation based reduction method as a contention resolution scheme in optical burst switching (OBS) network. In: Procedia Engineering (Vol. 30, pp. 1089–1096). <https://doi.org/10.1016/j.proeng.2012.01.967>
10. Espina Antolín F, Armendariz Silva J, Izal Azcárate M, Morató Osés D, Magaña Lizarrondo E (2009) Architecture and design of an OBS network model for simulation
11. Espina, F., Armendariz, J., Izal, M., Morató, D., & Magaña, E. (2009). Architecture and design of an OBS network model for simulation. *Jornadas de Ingeniería Telemática (JITEL)*.
12. Espina F, Armendariz J, García N, Morató D, Izal M, Magaña E (2010) OBS network model for OMNeT++ a performance evaluation. In: Proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques (pp 1–8)
13. Qiao C (2000) Labeled optical burst switching for IP integration over WDM. *IEEE Commun Mag* 38(9):104–114. <https://doi.org/10.1109/35.868149>
14. Gaona E, Gaona P, Montenegro C (2016) Optical switching devices in next generation networks. *Ingenium Revista de La Facultad de Ingeniería*, 14(27), 25. <https://doi.org/10.21500/01247492.2592>
15. Martín MÁT (2009) Evaluation of hybrid OBS/OCS network architectures. Barcelona
16. Argos CG (2008) Optical Burst Switching (OBS)
17. Peng L, Youn CH, Tang W, Qiao C (2012) A novel approach to optical switching for intradata-center networking. *J Lightwave Technol* 30(2):252–266. <https://doi.org/10.1109/JLT.2011.2180888>
18. Martínez JC, Veiga MF, González AS (2005) Burst assembly with proportional service differentiation in OBS networks
19. Farrington N, Forencich A, Porter G, Sun PC, Ford JE, Fainman Y, Vahdat A (2013) A multiport microsecond optical circuit switch for data center networking. *IEEE Photonics Technol Lett* 25(16):1589–1592. <https://doi.org/10.1109/LPT.2013.2270462>
20. Rostami A (2010) Traffic shaping for contention control in OBS networks
21. Puche W, Montoya G, Sierra J, Donoso Y (2009) Optical transport technologies: towards Optical Burst Switching (OBS). *Revista Investigaciones Aplicadas*, 2(2), 41–52. Retrieved from <https://revistas.upb.edu.co/index.php/investigacionesaplicadas/article/view/153>
22. Baldominos A, De Rada F, Saez Y (2018) DataCare: Big data analytics solution for intelligent healthcare management. *Int J Interact Multimed & Artif Intell*, 4(7)

# A Real-Time Arm-Worn Sensor-Based Human Fall Alert Notification Model for Efficient Daily Activity Recognition



Anurag De, C. Mugesh, Battula Lalitesh, and J. Joshua

**Abstract** One of the primary worries for the loved ones of elderly individuals is the accidental fall that can sometimes even lead them to death. Recent studies have shown that sensor-based methods in identifying falls mark a significant advancement in proactive fall detection, facilitating prompt assistance from concerned family members or caregivers to aid elderly individuals. Our primary objective revolves around the design and refinement of an arm-worn sensor-based fall detection method. Intelligently integrating accelerometer and gyroscope data into a wearable device, the model can efficiently analyze variations in different human daily activity movement patterns. This multifaceted approach enables our model to discern with precision between falls and activities of daily life (ADL). Through extensive experimentation and analysis, our results demonstrate the effectiveness of the proposed model in accurately identifying fall events while minimizing errors in detecting ADL. Achieving an impressive accuracy score of 98.66%, our novel approach proves to be more effective and beneficial in the realm of fall detection among elderly individuals compared to existing state-of-the-art methodologies.

**Keywords** Fall detection · Sensor-based approach · Accelerometer · Gyroscope · Activities of daily living (ADL)

## 1 Introduction

Annually, there are around 684,000 fatal falls, making it the second most common cause of unintentional injury deaths globally, after road traffic accidents. In 2019, nearly 3 million elderly individuals sought emergency care for fall-related injuries. The highest mortality rates due to falls are observed among adults aged 60 and above, resulting in nearly 32,000 older adult deaths each year [1]. As life expectancy rises and birth rates increase, this issue is observed to grow, with a larger global population aged

---

A. De (✉) · C. Mugesh · B. Lalitesh · J. Joshua

School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh, India

e-mail: [anurag.de111@gmail.com](mailto:anurag.de111@gmail.com)

65 and older. Research conducted in Canada, the United Kingdom, and the United States has revealed that the occurrence rates of falls among the elderly are roughly 30%, 20%, and 12%, respectively. The United States experienced a gradual increase in the percentage of mortality among older adults due to falls, rising from 31% in 2007 to 40% in 2016. This represents an annual growth rate of 3.0%. Meanwhile, several European countries have witnessed a significant annual rise of up to 20% in fall-related fatalities [2].

In addressing this issue, a multitude of approaches have arisen, prominently featuring sensors affixed directly to elderly individuals [3–5]. These sophisticated sensors are adept at discerning falls after they have transpired, promptly relaying essential information. Distinguishing between routine daily activities and accidental falls is a paramount aspect of these systems, ensuring accurate detection and response. Nonetheless, it's essential to acknowledge that current models and approaches exhibit certain constraints and shortcomings. The primary contributions of the proposed approach are:

- Our primary thrust is dedicated to developing an innovative body-worn sensor-based model that seamlessly integrates both accelerometer and gyroscope sensor readings in detecting a potential fall.
- It also notifies the caregivers about the potential fall, which helps them to promptly provide necessary care to the elderly people.
- This synergistic approach aims to not only enhance overall effectiveness but also comprehensively address the limitations inherent in each method, ultimately advancing the quality of care for the elderly.

The rest of the paper has been organized in the following order: A comprehensive review of the existing techniques for detecting human falls and activities of everyday life is provided in Sect. 2. The introduced fall detection approach is explained in Sect. 3. The outcomes of the experiment and the proposed system's performance evaluation are presented in Sect. 4, accompanied by a comparison with existing approaches. The paper's conclusion and potential for future improvements are described in Sect. 5.

## 2 Related Work

The research and literature that is currently available on various types of fall detection systems is compiled in this section.

Without requiring the user to wear any equipment, authors in [6] use commercial NIC equipment from Atheros to map wireless signal amplitude information to the human body's fall motion. To detect the fall, however, the transmitter and receiver must be present in the vicinity. A person's fall cannot be detected if they are outside the transmitter-receiver's range. Authors in [7] invented a mechanism for attaching it in the sole; if it detects a vertical sole, it declares it to be a fall. In this instance, the drawback is that the individual must wear the sole all the time to be monitored.

Hence, the mechanism is not a convenient solution for regular use. In [8], falls were detected by regularized logistic regression, and the accelerometer and gyroscope of smartphones were used to track movement. Information is stored on a cloud server and includes time, probability, location, weather, and activities prior to fall. In [9], the authors introduced a method that utilizes the accelerometer and gyroscope sensors of the Shimmer platform to transmit inertial signals to a computer. A compressed sensing technique is used in the system to minimize energy usage and reduce communicated data size. In a previous study by the authors in [10], they presented an algorithm framework for pre-alarming falls in the fractional domain. The data was then converted from time to the fractional domain. Using various machine learning algorithms, they achieved a greater pre-impact fall detection effect. However, while walking the sensors are entirely attached to the person's leg, which is a drawback. This method [11] determines the difference between an abrupt fall and a regular fall by measuring the instantaneous velocity and duration of falling. It incorporates several baseline spatial properties of an object in motion into the frame. To extract video frames depicting fall behavior and distinguish falls from ordinary activities, the authors in [12] introduced a stream-based set of criteria. Results are obtained by transferring the fall video frames recorded by detection to the MobileVGG network model. The primary limitation is its applicability solely to mobile devices. In [13], a deep learning-based non-intrusive HFD method is proposed. It uses an LSTM and CNN classifier to analyze camera footage and decide if a fall has taken place. A sensor, edge gateway, fog layer with LoRa communication, cloud layer, and application layer with user connectivity are the five steps that are modeled in [14]. Although this classification method is only applicable to photographs and not real-world circumstances, it has an impressive 98% accuracy rate in classifying positive and negative images utilizing parallel, vertical, and horizontal classification in fall detection.

### 3 Proposed Methodology

The proposed fall detection approach presents a body-worn sensor-based fall alert notification system. In our setup, we employed an accelerometer integrated with Node MCU to measure the speed of motion during a fall incident. To discern falls from routine daily activities, we established a threshold value; whenever the accelerometer surpasses this threshold, an alert email is automatically dispatched to a pre-registered email address through the IFTTT website [15]. The email notification includes a concise fall detection message.

#### 3.1 Arm-Worn Sensor-Based Approach

The arm-worn hardware device consists of a NodeMCU ESP8266 microcontroller, and a Wi-Fi module, to facilitate communication with the IFTTT website, which

sends notifications to the concerned person upon fall detection. Also, it includes the MPU6050 accelerometer, which provides crucial 3-axis data, and the gyroscope within the MPU6050, used for orientation determination. An overview of the device's operation is as follows.

**Step 1. Read sensor values:** Initially, the device reads sensor data from the MPU6050. It incorporates calibration values to adjust the raw sensor readings.

**Step 2. Calculate amplitude:** Using the accelerometer and gyroscope data, the device calculates the amplitude vector based on the accelerometer values as shown in Eq. (1).

$$\text{Amplitude (Acceleration change)}, A = \sqrt{ax^2 + ay^2 + az^2} \quad (1)$$

where the acceleration values along the  $x$ ,  $y$ , and  $z$  axes are represented, respectively, by the variables  $ax$ ,  $ay$ , and  $az$ . Equation (1) is used to find the overall acceleration experienced by a person in 3D space based on its acceleration along the  $x$ ,  $y$ , and  $z$ -axis. To find the overall acceleration magnitude, we should take the square root of the sum of squares of each component.

**Step 3. Trigger-1 activation:** If the accelerometer values surpass the lower threshold, trigger-1 is activated. It then checks for a change in accelerometer values beyond the higher threshold. If there's no change in the accelerometer value within this time, trigger-1 is deactivated.

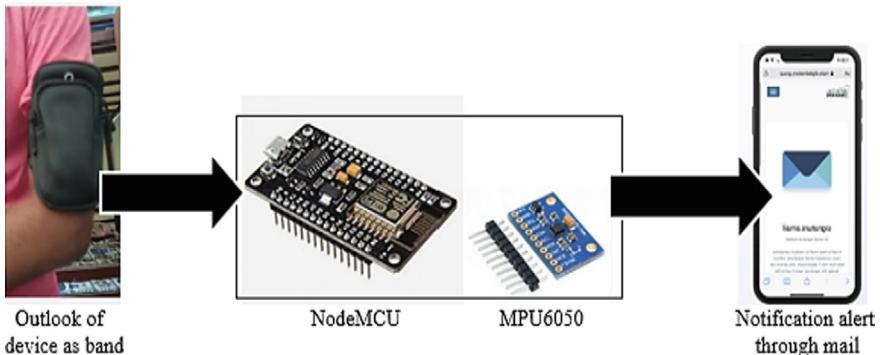
**Step 4. Trigger-2 activation:** If trigger-1 is activated, the device checks for values beyond the higher threshold. If this threshold is exceeded, trigger-2 is activated.

**Step 5. Trigger-3 activation:** It calculates the change in orientation as shown in Eq. (2) using gyroscope values. If there's a sudden change in orientation, trigger-3 is activated; otherwise, trigger-2 is deactivated.

$$\text{Orientation (Angle change)}, O = \sqrt{gx^2 + gy^2 + gz^2} \quad (2)$$

where  $gx$ ,  $gy$ , and  $gz$  represent the gyroscope values measured along the  $x$ ,  $y$ , and  $z$ -axis respectively. Equation (2) calculates the orientation change based on gyroscope measurements along the  $x$ ,  $y$ , and  $z$ -axis. It is used for orientation tracking. It represents how quickly the orientation of the object is changing in 3D space, regardless of the specific direction of the change along each axis. To find the overall orientation change magnitude, we should take the square root of the sum of the squares of each component. If trigger-3 is activated, the device waits for 10 s to confirm whether a fall has genuinely occurred by checking if the orientation value remains unchanged. If it remains the same, the fall detection signal is activated, and an event is sent through IFTTT.

**Step 6. Notification alert:** To transmit the signal, IFTTT utilizes the Wi-Fi host, which requires the username, password, and private key of the website. If the host successfully connects to the client's HTTP port, the connection is established. The device then creates a URL for the request using the private key and sends the message "*Fall Detected*" to notify the designated person. After sending the mail, trigger-3 is



**Fig. 1** System architecture of the arm-worn sensor-based approach

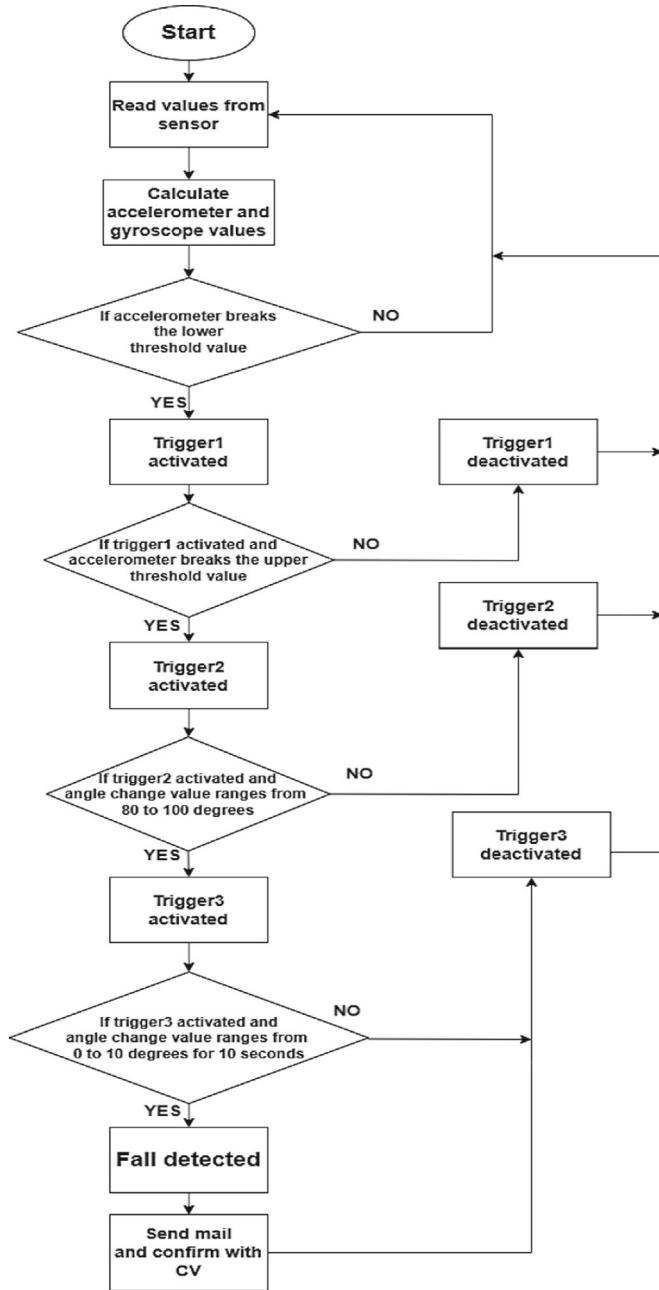
deactivated. The arm-worn hardware device depicted in Fig. 1 functions as the main means of verifying a detected fall. However, computer vision techniques can also be used for further verification and validation of the detected fall movement. If the hardware component initiates a fall detection notification, it indicates that a potential fall has been identified. Wearing this device on the arm provides a key advantage since it guarantees the individual's comfort without any disturbance.

Figure 2 presents an elaborate flowchart illustrating the workflow of the body-worn sensor-based system, describing the sequential phases involved in the process of detecting falls. This method depends on the observation of alterations in acceleration and orientation values to initiate a reaction.

## 4 Experiment and Results

This section discusses the results which are obtained by our proposed model and gives the performance evaluation along with a comparison analysis of various existing state-of-the art fall detection approaches. The status of each trigger tracked by the body-worn sensor device when an event (simulated by volunteers) is in progress is shown in Table 1. This comprehensive algorithm enables the system to determine whether the observed activity corresponds to a fall or if it is simply an ADL.

For the first activity in Table 1, which is standing, trigger-1 is activated and checks for trigger-2 and trigger-3 threshold values. As there is very little variation in acceleration trigger-2 is deactivated and it does not check for trigger 3 and classify it as an ADL. In the case of sitting activity, there is a change in acceleration while sitting quickly and a change in the gyroscope readings because the angle also changes accordingly. So, all three triggers are activated but neither of them reached the threshold limit. Also, there is significant body movement after the activity. So, it is also considered as an ADL. Similarly, for bending posture, trigger-1 is activated and checks for remaining trigger threshold values. There is a change in angle but



**Fig. 2** Algorithm of the arm-worn sensor-based approach

**Table 1** Activity recognition results of a volunteer by the arm-worn sensor-based approach

Activity	Trigger-1	Trigger-1 Status	Trigger-2	Trigger-2 Status	Trigger-3	Trigger-3 Status	Movement Within threshold limit (10 s)	Decision
Standing	1	Activated	7	Deactivated	–	Deactivated	–	ADL
Sitting	1	Activated	17	Activated	201	Activated	Yes	ADL
Bending	0	Activated	12	Deactivated	–	Deactivated	–	ADL
Crouching	1	Activated	24	Activated	248	Activated	Yes	ADL
Falling	2	Activated	37	Activated	276	Activated	No	Fall

**Table 2** Confusion matrix of the proposed fall detection model

	Actual Fall	Actual ADL
Detected Fall	35 (TP)	01 (FP)
Detected ADL	00 (FN)	39 (TN)

not greater than the threshold limit. So, it is also classified as an ADL. In the fourth activity, i.e. crouching, all three triggers get activated but neither of them reaches its maximum value to declare it as a fall. Also, there is body movement of the subject after crouching down. Hence, it is not considered a fall and declared as an ADL. However, in the fifth simulation, a sudden variation in the accelerometer and gyroscope values is observed, and simultaneously there is no movement after the activity up to a certain duration. Consequently, it is classified as a potential fall behavior and an alert notification is sent for necessary care.

#### 4.1 Performance Evaluation

Table 2 displays the confusion matrix for the suggested model. It shows that all the 35 volunteer-simulated fall movements are identified as a fall with accuracy. 39 projected activities of daily living (ADLs) are identified from the 40 simulated daily life activities, while one action is expected to be a fall. The quantitative performance of the proposed model based on specificity, sensitivity, accuracy, and precision is shown in Table 3. A detailed comparison between our suggested method and state-of-the-art fall detection models currently in use can be seen in Table 4.

**Table 3** Quantitative performance of the proposed approach

Approach used	Specificity (In %)	Sensitivity (In %)	Precision (In %)	Accuracy (In %)
Threshold-based	97.5	100	97.22	98.66

**Table 4** Comparison with existing state-of-the-art fall detection approaches

Approach	Acc score	Drawbacks	Overcome by the proposed approach
Duan et al. [3]	92.3%	Falls detected only in the transmitter range	The sensors will help in this regard
Zitouni et al. [4]	-	Sole worn constantly	Comfortably worn on hand like a band
Chaithanya et al. [5]	95.9%	Elderly people may not carry smartphones	A separate device without their phones
Our Approach	98.66%	-	-

## 5 Conclusion and Future Directions

Simulating an effective fall alarm notification system with a body-worn device that is easy to use, the proposed technique has addressed the urgent problem of unintentional falls in the elderly population. By monitoring the readings of accelerometer and gyroscope sensors embedded to the armband, our model can accurately detect probable falls while minimizing inaccuracies in identifying everyday activities. The proposed fall detection model simulates everyday falls and fall-like actions through young volunteers. It is extremely difficult, and often even deadly, to utilize older people in simulation. The proposed fall detection model achieves a satisfactory accuracy score of 98.66%. It outperforms all compared cutting-edge methods in effectively and accurately identifying falls and notifying the concerned personnel when an emergency arises. To improve our suggested method in the future, we plan to incorporate both traditional machine learning and modern deep learning techniques. Integration of computer vision-based techniques for enhancing the overall robustness of the system can also be considered as a future scope.

## References

1. WHO, Falls, <https://www.who.int/news-room/factsheets/detail/falls>. Last accessed 2024/05/05
2. Soomar SM, Dhalla Z (2023) Injuries and outcomes resulting due to falls in elderly patients presenting to the emergency department of a tertiary care hospital—a cohort study. BMC Emergency Medicine 23, (2023)
3. Li S (2023) Fall Detection with Wrist-Worn watch by observations in statistics of acceleration. IEEE Access 11:19567–19578. <https://doi.org/10.1109/ACCESS.2023.3249191>
4. Lin H-C, Chen M-J, Lee C-H, Kung L-C, Huang J-T (2023) Fall recognition based on an IMU wearable device and fall verification through a smart speaker and the IoT. Sensors 23(12). <https://doi.org/10.3390/s23125472>
5. Subramaniam S, Faisal AI, Deen MJ (2022) Wearable sensor systems for fall risk assessment: a review. Front Digit Health 4. <https://doi.org/10.3389/fdgh.2022.921506>
6. Yu D, Hao Z, Dang Z, Xu H (2019) KS-FALL: Indoor human fall detection method under 5GHz wireless signals. In: IOP Conference Series: Materials Science and Engineering 569(3). <https://doi.org/10.1088/1757-899X/569/3/032068>
7. Zitouni M, Pan Q, Brulin D, Campo E (2019) Design of a smart sole with advanced fall detection algorithm. J Sens Technol 9(4). <https://doi.org/10.4236/jst.2019.94007>
8. Harari Y, Shawen N, Chaithanya KM, Albert MV, Kording KP, Jayaraman A (2021) A smartphone-based online system for fall detection with alert notifications and contextual information of real-life falls. J Neuro Eng Rehabil 18 (124). <https://doi.org/10.1186/s12984-021-00918-z>
9. Oussama K, Ramzan N, Ghanem K et al (2019) Fall detection and human activity classification using wearable sensors and compressed sensing. J Ambient Intell Humaniz Comput 11:349–361. <https://doi.org/10.1007/s12652-019-01214-4>
10. Ning L, Zhang D, Su Z, Wang T (2021) Preimpact fall detection for elderly based on fractional domain. Math Probl Eng 2021. <https://doi.org/10.1155/2021/6661034>
11. De A, Saha A, Kumar P (2022) Fall detection approach based on combined displacement of spatial features for intelligent indoor surveillance. Multimed Tools Appl 81:5113–5136. <https://doi.org/10.1007/s11042-021-11646-w>

12. Han Q, Zhao H, Min W et al (2020) A two-stream approach to fall detection with MobileVGG. *IEEE Access* 8:17556–17566. <https://doi.org/10.1109/ACCESS.2019.2962778>
13. Alam E, Sufian A, Dutta P, Marco L (2022) Vision-based human fall detection systems using deep learning: A review. *Computers in Biology and Medicine* 146. <https://doi.org/10.1016/j.combiomed.2022.105626>
14. Vimal S, Robinson YH, Kadry S, et al. (2021) IoT based smart health monitoring with CNN using edge computing. *J Internet Technol* 22 (1)
15. IFTTT website, <https://ifttt.com/about>, Last accessed 2024/05/05

# Mobile Applications in Electronic-Healthcare: A Case Study for Bangladesh



Jarin Tasnim , Shamim Forhad , Sunjida Mushfiq Nova ,  
Khandakar Kamrul Hasan , S. M. Nahidul Islam ,  
Samia Binta Hassan , Mohammad Ashiqur Noor ,  
and Abdul Hasib Siddique

**Abstract** Bangladesh's healthcare system faces major problems, like unequal access, high costs for patients, not enough skilled healthcare workers, and more cases of non-communicable diseases (NCDs). The system focuses too much on hospitals and doctors, which makes it harder for people, especially in rural areas, to get quality care. Using healthcare apps can help solve these problems by making care more accessible, affordable, and better overall. This research represents how healthcare applications are currently used in Bangladesh and their effects on society. It aims to show the benefits of using mobile healthcare apps, focusing on key examples in Bangladesh. The study also explores the e-healthcare environment in the country, identifying obstacles that prevent these apps from being fully effective and suggesting ways to overcome them. The findings highlight the positive social impact of mobile apps in Bangladesh, including better access to healthcare, improved patient outcomes, and a shift towards more personal independence in the healthcare system.

**Keywords** Healthcare-Applications · Telemedicine · IoT · Bangladesh · Real-Time monitoring · Teleconsultation

## 1 Introduction

In the contemporary healthcare landscape, integrating technology has become paramount in addressing the evolving needs of patients and healthcare providers. Among the various technological innovations, mobile applications stand out as powerful tools capable of revolutionizing healthcare delivery. A significant concern is the disparity in healthcare access and affordability, particularly pronounced in

---

J. Tasnim · S. Forhad ( ) · S. M. Nova · K. K. Hasan · S. M. N. Islam · S. B. Hassan ·  
M. A. Noor · A. H. Siddique  
University of Scholars, 40 Kemal Ataturk Ave, Dhaka 1213, Bangladesh  
e-mail: [nrshamimforhad@gmail.com](mailto:nrshamimforhad@gmail.com)

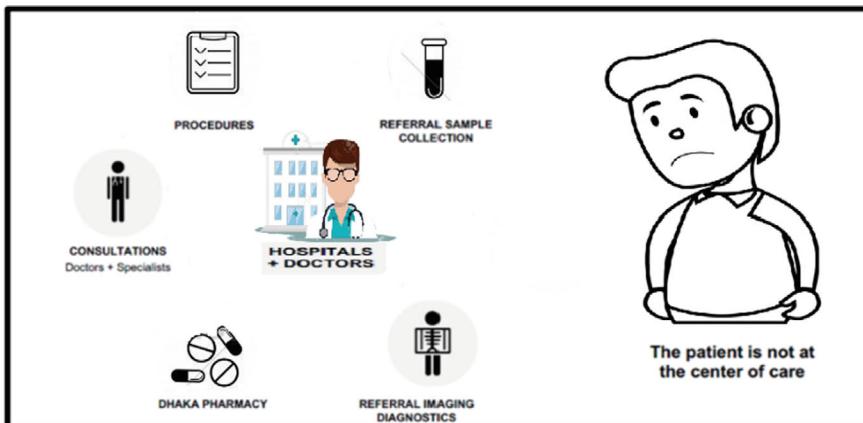
rural and underserved regions [1]. The financial burden of out-of-pocket expenses for medical services remains high, often leading to economic strain on families [2].

The rising burden of non-communicable diseases (NCDs) such as heart disease, diabetes, and cancer present a formidable challenge, with the healthcare system currently underprepared in terms of infrastructure to manage this growing epidemic [3]. Moreover, the financing model of the healthcare system, heavily reliant on out-of-pocket payments, limits access to necessary healthcare services for many, highlighting the urgent need for more robust health insurance schemes and increased government funding [4].

Another major problem which is visible at Fig. 1, that is the healthcare system in regions like Bangladesh tends to be more focused on hospitals and doctors rather than being patient centric. This approach often prioritizes the availability of healthcare professionals and facilities over the individual needs and preferences of patients, leading to a care process where patients may feel like passive recipients [5]. Leveraging technology can also play a key role in making healthcare more accessible and responsive to patient needs [6] (Fig. 2).

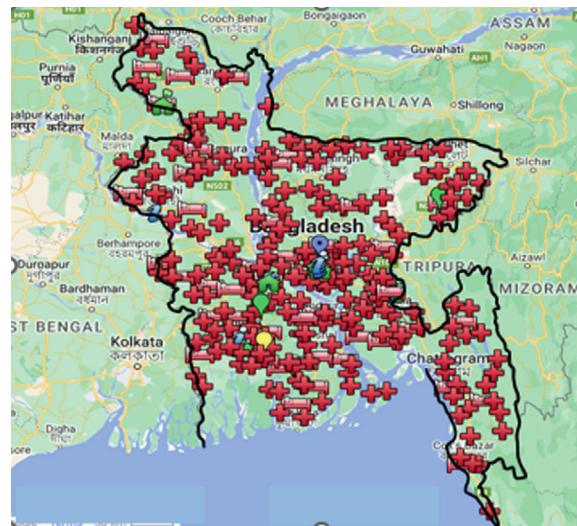
In Table 1 the approximate data by the Bangladesh Bureau of Statistics and the Ministry of Health and Family Welfare are mentioned. population of Bangladesh is approximately 165 million (as of 2023 estimates), the area is 147,570 square kilometers, total hospitals are around 5,816 (including public and private), hospitals in each division may vary, as this will be a rough estimate based on available data, ratio of population to hospitals were calculated based on available population and hospital numbers, ratio of doctor to patient has been a challenge to calculate, with an estimated ratio previously reported as roughly 1 doctor to every 2,000 patients [7–11].

According to the World Health Organization (WHO), mHealth can be described as the application of mobile devices, including mobile phones, patient monitoring devices, personal digital assistants (PDAs), and various other wireless gadgets, to



**Fig. 1** Traditional healthcare system in Bangladesh

**Fig. 2** Healthcare infrastructure in Bangladesh



**Table 1** Healthcare infrastructure and demographic analysis by division in Bangladesh

Division	Population (Est.) in Million	Area (sq km)	Number of Hospitals (Est.)	The ratio of Population to Hospitals	The ratio of Doctor to Patient (Approx.)
Dhaka	20 +	20,593	2,000 +	10,000: 1	1: 1847
Chattogram	30 +	33,771	1,000 +	30,000: 1	1: 1847
Rajshahi	20 +	18,174	500 +	40,000: 1	1: 1847
Khulna	15 +	22,272	400 +	37,500: 1	1: 1847
Barishal	8 +	13,297	300 +	26,666: 1	1: 1847
Sylhet	10 +	12,596	300 +	33,333: 1	1: 1847
Rangpur	15 +	16,184	500 +	30,000: 1	1: 1847
Mymensingh	15 +	10,584	400 +	37,500: 1	1: 1847
Overall	165 +	147,570	5,816	28,375: 1	1: 1847

support public health as well as medical standards [13]. Despite the burgeoning presence of these digital health solutions, the country faces unique challenges that could impede their full-scale implementation and effectiveness [14]. The adoption of mobile healthcare applications in Bangladesh has the potential to significantly improve healthcare service delivery, particularly in rural and underserved areas where healthcare resources are scarce [15].

The value of mobile applications in e-healthcare is extensive and has profound effects on how healthcare services are being delivered [17, 18]

This paper explores the current state and consequences of healthcare applications in Bangladeshi society, emphasizing their impact on enhancing healthcare accessibility, patient outcomes, and technological utilization. There is a notable emphasis on matters about the security of medical data, which has become a significant concern due to recent instances of illegal public access to confidential health information [19]. The presence of these vulnerabilities highlights the pressing necessity for strong data protection procedures to ensure the preservation of patient privacy and the establishment of confidence in digital healthcare solutions [20]. The motivation of this research is to assemble the scattered mobile apps available in the market and perform a comparative analysis so that the mass population is aware of the healthcare apps and their performance. Besides that, this case study aims to explore the role of mobile applications in augmenting observability in Bangladesh's healthcare sector. Observability, in this context, refers to the ability to gather real-time data, monitor health indicators, track medical histories, and facilitate timely interventions, ultimately leading to improved health outcomes and efficient resource allocation.

The paper includes Current Trends in Healthcare Applications in Bangladesh, Notable Mobile Applications in Bangladeshi Electronic-Healthcare from Past to Present, Challenges And Considerations In E-Health Mobile App Implementations in Context to Bangladesh, Future Scope for Mobile Apps in Electronic-Healthcare, Conclusion etc. sections to fulfill the research goal.

## 2 Current Trends in Healthcare Applications in Bangladesh

Trending Healthcare applications in Bangladesh, including Praava Health, Maya, Arogga, BanglaMeds, Olwel, AmarLab, LifeSpring, Pulse Healthcare Services, Moner Bondhu, and Jolpie, have each carved out their niche in the market. These companies are not just addressing the traditional healthcare delivery challenges but are also pioneering in leveraging technology to cater to the specific needs of the Bangladeshi population. From telemedicine and digital pharmacies to mental health support and beyond, each app has its unique proposition [21].

For instance, Praava Health is blending the convenience of telemedicine with the assurance of physical clinics, offering a comprehensive healthcare experience [21]. Meanwhile, Maya extends its services beyond healthcare to include psycho-social and legal advice, showcasing the multifaceted needs of health app users [22]. Arogga and BanglaMeds are streamlining the process of medication delivery and healthcare procurement, making it easier for individuals to receive their prescriptions and other health-related products without the need to leave their homes [23, 24]. Similarly, Olwel, DocTime, and AmarLab focus on bringing healthcare services directly to the patient's doorstep, eliminating the geographical and logistical barriers that often impede access to quality healthcare [25, 26].

LifeSpring and Moner Bondhu are addressing the critical but often overlooked aspect of mental health, providing platforms for counseling and psychological support [27, 28]. DocTime leverages the power of telemedicine to offer virtual

consultations, making healthcare more accessible, especially in remote areas or for individuals with mobility challenges [29]. By filling gaps in the healthcare system, these apps play a pivotal role in enhancing patient care and facilitating better health outcomes across the country.

### **3 Notable Mobile Applications in Bangladeshi Electronic-Healthcare from Past to Present**

#### ***3.1 Praava Health***

Praava Health has revolutionized healthcare with quality services, including diagnostics and pharmacy, addressing the country's healthcare deficiencies. Despite its innovative approach, Praava faced challenges in fundraising. This highlights the struggle innovative healthcare solutions encounter in securing financial backing, despite demonstrating potential to significantly improve patient care and healthcare accessibility.

#### ***3.2 Maya Healthcare Solutions***

Maya provides multilingual consultations, efficiently handling most inquiries through its AI technology. Catering to over 15 million users, the platform maintains anonymity for users while offering expert advice and facilitating medical purchases. Following a recent financial boost, Maya is on a path to further integrating AI into healthcare, aiming to reach underserved populations, especially in rural areas.

#### ***3.3 JOTNO***

JOTNO is a digital health service platform in Bangladesh designed to improve healthcare accessibility and efficiency for both doctors and patients. It offers features such as video consultations, appointment scheduling, and health packages. For patients, JOTNO simplifies the process of seeking medical advice, appointments, prescriptions, and health insurance purchases.

### ***3.4 Arogga***

Quite similar to Tonic and Doctorola it also provides an online platform for healthcare that offers telemedicine services and puts people in touch with doctors. Besides the common features, it also provides information, advice, and reminders for taking medications as prescribed.

### ***3.5 AmarDoctor***

This is an AI-based health assistant to find doctors according to patients' needs by categorizing the symptoms. This app offers services like finding a doctor, checking symptoms, booking a call with a doctor, and describing different medicines and information of different diseases through the app.

### ***3.6 MedEasy***

Patients may schedule appointments, look up doctors by expertise from 23 different specialized areas, and communicate with medical experts via video calls or chat. For the convenience of users, the app also offers access to medicines and medical reports which is very important for future reference for doctors' consultation.

### ***3.7 DaktarBhai***

Daktarbhai combines technology and healthcare to provide online medical services, including appointment booking, electronic health record management, and access to health resources and discounts. Daktarbhai aims to simplify healthcare access by offering free electronic health records, a health services directory, and medication reminders.

### ***3.8 Sebaghar***

Sebaghar offers seamless video consultations with doctors, online prescription services for efficient medication management, and a system for scheduling prompt appointments with healthcare professionals. The platform is designed to be user-friendly, making healthcare accessible even to those in remote areas or with limited time.

### **3.9 Doctorola**

Users may use the app to schedule appointments and search for doctors based on specialization, availability, and region. A comprehensive range of healthcare services, including doctor consultations, diagnostic procedures, and drug delivery are being offered through the platform.

### **3.10 Amar Lab**

The telemedicine app Amar Lab is largely focused on providing diagnostic services. Users may schedule lab tests and obtain digital results. Users of the program may interact with doctors and get guidance depending on the outcomes of their tests.

In this research we have done comparison between the features that the apps are offering which has been showed in Table 2 and it finds that the Doctorola application allows users to schedule appointments with doctors and provides access to health information, but it lacks features for tracking your health status or managing your medications.

Arogga, other apps offer the option of health monitoring and tracking, but they do not include medication management, telemedicine, or doctor appointment services. While Amar Lab and Digital Healthcare merely provide the option of telemedicine and virtual consultation. Praava health is offering necessary features, however depending.

on customer demands more options can be improved. Most of the users desired to discover facilities that combine all available possibilities into a single app and fulfill their needs within one click. These apps empower patients by providing health information, self-care options, and tailored resources, enhancing patient engagement and healthcare outcomes.

Market research is paramount for understanding the demands and preferences of the local audience, allowing software to be adapted to the language, culture, and healthcare system of Bangladesh. Continuous enhancement of the app, guided by user input and evolving market conditions, is crucial. Through a comprehensive study of available apps in the market, a model incorporating all desired features has been proposed. Also, integrating extra functionalities such as emergency calls to nearby hospitals and primary first aid support can transform the healthcare app into a profitable business venture that also contributes significantly to society.

**Table 2** Comparison between e-health apps used in Bangladesh

E-Healthcare apps (BD)	Doctor appointment & Communication	Telemedicine & virtual consultation	Health Monitoring	Medication management	Information and education resources
Doctorola	✓	✓	✗	✗	✓
Maya	✓	✓	✗	✗	✓
Aroga	✓	✗	✓	✗	✗
AmarDoctor	✓	✗	✗	✗	✗
HealthifyMe	✗	✗	✓	✗	✗
Fitbit	✗	✗	✓	✗	✗
MiFit	✗	✗	✓	✗	✗
Samsung health	✗	✗	✓	✗	✗
MyfitnessPal	✗	✗	✓	✗	✗
GoogleFit	✗	✗	✓	✗	✗
Amar Lab	✗	✓	✗	✗	✗
MediSure	✗	✗	✗	✓	✗
Medikare	✗	✗	✗	✓	✗
RemindMyMeds	✗	✗	✗	✓	✗
HealthKeeper	✗	✗	✓	✗	✗
Praava health	✓	✓	✓	✓	✓
Jotno	✗	✗	✗	✗	✗
MedEasy	✗	✗	✗	✓	✗
Daktar Bhai	✓	✓	✗	✓	✓
Sebaghar	✗	✓	✗	✗	✓

## 4 Challenges and Considerations in E-Health Mobile App Implementations in Context to Bangladesh

### 4.1 Limited Internet Connectivity

Bangladesh still has issues with internet accessibility, especially in rural regions. Bangladesh has been involved in submarine cable projects to enhance international connectivity and reduce dependency on costly satellite bandwidth. Initiatives are also underway to set up community internet centers in rural areas to provide access to internet services for underserved populations, hence it is strongly believed that by 2030 this issue would be resolved [30].

## ***4.2 Digital Literacy***

Another critical issue which is limiting the progress of e-healthcare apps is digital literacy. Looking at the huge population which is over the age of 60 it could have been a timely solution to combat the challenge of lack of Hospitals or doctors. Many people, especially those who live in rural regions and elderly persons, may not be very familiar with cellphones and mobile apps [31]. To bridge the digital literacy gap, organizations must provide user-friendly interfaces, deliver clear instructions, and provide support.

## ***4.3 Education Among the Common Mass***

As a large population of our country can't really navigate through this complex app hence almost all the apps failed to popularize them. To address these challenges, the app can offer step-by-step guidance, short instructional videos, and comprehensive manuals with troubleshooting tips. Additionally, user engagement can be enhanced through quizzes, challenges, and rewards. Importantly, gathering user feedback for continuous improvement is key.

## ***4.4 Language and Localization***

In order to maintain successful communication and user engagement, Bangla language support and consideration of language preferences are crucial in e-health mobile apps. Highlighting culturally sensitive design and content localization is crucial for e-health apps in Bangladesh. This involves tailoring the app's interface, content, and features to align with the cultural norms and preferences of the Bangladeshi population. Moreover, considering factors like religious observations, healthcare beliefs, and social customs can further enhance the app's relevance and effectiveness [32].

## ***4.5 Health Equity and Accessibility***

Bangladesh has a huge population who are living under extreme financial stress. In order to incorporate all sections of the population, particularly underprivileged groups, must be able to use and benefit from e-health mobile apps. To guarantee equal access to healthcare services, efforts should be made to close the digital gap, consider affordability and accessibility issues, and aid those with impairments.

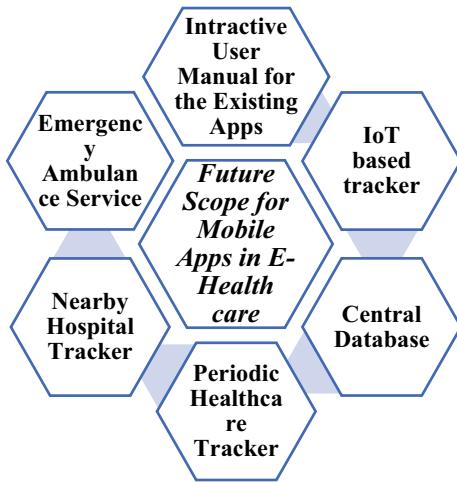
## 5 Future Scope for Mobile Apps in Electronic-Healthcare

There are enormous existing apps in the e healthcare system of Bangladesh but there are some features that can be merged with the current app's features so that it be implement for better and efficient health care system. Some of the proposed approaches and features have been stated below:

- **Improved Patient Monitoring and Remote Care:** Integrating IoT trackers into e-healthcare apps can enable real-time monitoring of patient's vital signs, medication adherence, and overall health status. This feature will facilitate remote care management, particularly for patients in rural or underserved areas where access to healthcare facilities is limited.
- **Efficient Resource Allocation:** Centralizing patient data in a secure and accessible central database will streamline information sharing and collaboration among healthcare providers. This can lead to reduced healthcare costs, minimized duplication of tests, and improved efficiency in delivering healthcare services across different regions of Bangladesh.
- **Facilitation of Telemedicine and Teleconsultation:** The integration of IoT trackers and central databases in e-healthcare apps can facilitate telemedicine and teleconsultation services, enabling patients to receive virtual healthcare consultations from qualified providers. Bangladesh can overcome barriers to healthcare access, reduce healthcare disparities, and enhance the overall quality of care delivery.

The future scope for electronic-healthcare in Bangladesh can be visualized in Fig. 3 which is promising with the integration of IoT trackers and central databases into mobile health apps. All the addressed feature's importance in the existing healthcare system have been presented in this section. An interactive User Manual for the Existing Apps can be beneficiary for the masses to guide them in using all these apps, where an IoT based tracker will help the people to integrate the apps with locations as well as with the central database of the patients. A periodic health-care tracker is important for the patients to make them conscious for periodic health checkups and nearby hospitals and emergency notification is a most required addition that can be added to the apps for emergency service for the patients. By harnessing the power of technology to enhance patient monitoring, resource allocation, disease surveillance, community health outreach, and telemedicine services, Bangladesh can achieve significant advancements in healthcare delivery, ultimately improving health outcomes and promoting equitable access to healthcare services for all its citizens.

**Fig. 3** Future scope for mobile apps in E-Health care



## 6 Conclusion

In conclusion, leveraging mobile applications for e-healthcare has the potential to transform Bangladesh's healthcare sector by addressing some of its most critical challenges. The widespread availability and use of smartphones offer a practical and accessible platform for delivering healthcare services, enhancing patient care, and improving healthcare efficiency. Mobile applications can bridge the gap between healthcare providers and patients, especially in rural areas where access to healthcare is often limited. These apps can facilitate remote consultations, real-time health monitoring, and prompt medical advice, thereby reducing the need for physical visits and associated costs. For mobile apps to be successfully integrated into Bangladesh's e-healthcare system, several factors must be carefully considered. Understanding the local market is crucial, as it involves recognizing the specific health needs and preferences of the Bangladeshi population. User preferences must be taken into account to ensure that the apps are user-friendly, culturally relevant, and cater to the varying levels of digital literacy across different demographics. Additionally, infrastructural constraints such as internet connectivity and smartphone penetration rates need to be addressed to ensure broad accessibility and usability. Regular updates and enhancements based on user feedback and technological advancements are essential to maintain the apps' relevance and effectiveness. Efficient marketing strategies are necessary to raise awareness and encourage widespread adoption among both healthcare providers and patients. Furthermore, adherence to legal standards and regulatory requirements is vital to ensure data privacy, security, and compliance with healthcare regulations. Improved access to healthcare services through mobile apps can significantly enhance health outcomes, reduce disparities in healthcare quality and access, and ultimately contribute to a more equitable healthcare system. By closing the gaps in healthcare quality and access, mobile applications can play a pivotal role

in advancing the overall well-being of the Bangladeshi population, paving the way for a healthier and more prosperous future.

## References

1. Asia Pacific Observatory on Health Systems and Policies, “Bangladesh Health System Review,” World Health Organization Regional Office for South-East Asia, (2015). Available: <https://apo.who.int/publications/i/item/9789290617051>. (Accessed: 31-Mar-2024)
2. LightCastle Partners, “Healthcare Sector in Bangladesh: Into the Future,” LightCastle Partners, (2022). Available: [www.lightcastlebd.com](http://www.lightcastlebd.com). (Accessed: 31-Mar-2024)
3. Biswas T, Azzopardi P, Anwar SN et al (2022) Assuring Bangladesh’s future: non-communicable disease risk factors among the adolescents and the existing policy responses. J Health Popul Nutr 41:22. <https://doi.org/10.1186/s41043-022-00294-x>
4. Biswas T, Pervin S, Tanim MIA et al (2017) Bangladesh policy on prevention and control of non-communicable diseases: a policy analysis. BMC Public Health 17:582. <https://doi.org/10.1186/s12889-017-4494-2>
5. Asia Pacific Observatory on Health Systems and Policies, “Bangladesh Health System Review,” WHO Regional Office for South-East Asia, (2015). Available: <https://apo.who.int/publications/i/item/9789290617051>. (Accessed: Mar. 31, 2024)
6. Xesfingi S, Vozikis A (2016) Patient satisfaction with the healthcare system: Assessing the impact of socio-economic and healthcare provision factors. BMC Health Serv Res 16:94. <https://doi.org/10.1186/s12913-016-1327-4>
7. Doctors Gang, “Doctor Population Ratio in Bangladesh,” Doctors Gang, Available: <https://www.doctorsgang.com/doctor-population-ratio-in-bangladesh-doctors-gang/>. (Accessed: Mar. 31, 2024)
8. Asia Pacific Observatory on Health Systems and Policies, “Bangladesh Health System Review,” WHO Regional Office for South-East Asia, (2015). Available: <https://apo.who.int/publications/i/item/9789290617051>. (Accessed: 31-Mar-2024)
9. World Health Organization, “WHO Bangladesh Country Cooperation Strategy: 2020–2024,” WHO, (2023). Available: <https://www.who.int/publications/i/item/9789290209478>. (Accessed: 31-Mar-2024)
10. World Health Organization, “Environmental Health Bangladesh 2023 Country Profile,” WHO, (2023). Available: <https://www.who.int/publications/m/item/environmental-health-bdg-2023-country-profile>. (Accessed: 31-Mar-2024)
11. World Health Organization, “Immunization Bangladesh 2023 Country Profile,” WHO, (2023). Available: <https://www.who.int/publications/m/item/immunization-bangladesh-2023-country-profile>. (Accessed: 31-Mar-2024)
12. Anwar Islam, Tuhin Biswas (2014) Health System in Bangladesh: Challenges and Opportunities. Am J Health Res, 2(6), 366–374. <https://doi.org/10.11648/j.ajhr.20140206.18>
13. WHO Global Observatory for eHealth (2011) mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organ. <https://iris.who.int/handle/10665/44607> (Accessed: 03 March, 2024).
14. Mumtaz H, Riaz MH, Wajid H, Saqib M, Zeeshan MH, Khan SE, Chauhan YR, Sohail H, Vohra LI (2023) Current challenges and potential solutions to the use of digital health technologies in evidence generation: a narrative review. Front Digit Health. 28(5):1203945. <https://doi.org/10.3389/fdghth.2023.1203945.PMID:37840685;PMCID:PMC10568450>
15. Hasan U, Nazrul Islam M, Tajmim Anuva S, Rahman Tahmid A (2021) User-Centred Design-Based Privacy and Security Framework for Developing Mobile Health Applications. In: Uddin MS, Bansal JC (eds). In: Proceedings of International Joint Conference on Advances in Computational Intelligence. Algorithms for Intelligent Systems. Springer, Singapore. [https://doi.org/10.1007/978-981-16-0586-4\\_17](https://doi.org/10.1007/978-981-16-0586-4_17)

16. Neumeyer X, Santos SC, Morris MH (2021) Overcoming barriers to technology adoption when fostering entrepreneurship among the poor: the role of technology and digital literacy. *IEEE Trans Eng Manage* 68(6):1605–1618. <https://doi.org/10.1109/TEM.2020.2989740>
17. Kher RK, Mistry A (2022) e-Healthcare challenges: scenario in rural regions of South Asia. In: Kher RK, et al. IoT applications for healthcare systems. EAI/Springer Innovations in Communication and Computing. Springer, Cham. [https://doi.org/10.1007/978-3-030-91096-9\\_6](https://doi.org/10.1007/978-3-030-91096-9_6)
18. Mostafa R, Ehsanur Rahman GMA, Hasan GM, Kabir A, Rahman A, Ashik S (2010) Proposed deployments to provide E-healthcare in Bangladesh: Urban and rural perspectives. In: The 12th IEEE International Conference on e-Health Networking, Applications and Services, Lyon, France, 2010, pp 361–366, <https://doi.org/10.1109/HEALTH.2010.5556542>
19. Arkorful VE, Shuliang Z, Muhideen S, Basiru I, Hammond A (2020) An empirical investigation of health practitioners technology adoption: the mediating role of electronic health. *Int J Public Adm* 43(12):1013–1028. <https://doi.org/10.1080/01900692.2019.1664569>
20. Javad Pool, Saeed Akhlaghpour, Farhad Fatehi, Andrew Burton-Jones (2024) A systematic analysis of failures in protecting personal health data: A scoping review. *Int J Inf Manag.* 74, 102719, ISSN 0268-4012, <https://doi.org/10.1016/j.ijinfomgt.2023.102719>
21. Chu, Michael (2021) Praava health: a new model for Bangladesh. Harvard Business School Case 322–067, November 2021. (Revised December 2021)
22. Marzan Jui U (2021) Maya: An AI-powered start-up that plans to become the go-to name in healthcare. The Business Standard. Available: <https://www.tbsnews.net/features/panorama/maya-ai-powered-start-plans-become-go-name-healthcare-314215>. (Accessed: Mar. 27, 2024)
23. The Daily Star Youth (2024) Jotno: facilitating healthcare to your doorstep. The Daily Star. Available: <https://www.thedailystar.net/star-youth/news/jotno-facilitating-health-care-your-doorstep-1873210>. (Accessed: Mar. 31, 2024)
24. Jahan Khan J (2020) Online medicine delivery services in Dhaka during quarantine. The Daily Star. Available: <https://www.thedailystar.net/toggle/news/online-medicine-delivery-services-dhaka-during-quarantine-1896028>. (Accessed: Mar. 27, 2024)
25. Dynamic Solution Innovators, “Olwel,” Dynamic Solution Innovators. Available: <https://www.dsinnovators.com/projects/olwel>. (Accessed: Mar. 27, 2024)
26. Marzan Jui U (2020) AmarLab: Diagnosis at your doorstep. The Business Standard. Available: <https://www.tbsnews.net/bangladesh/health/amarlab-diagnosis-your-doorstep-48229>. (Accessed: Mar. 27, 2024)
27. Sayedul Ashraf, Yahia Md Amin, Md Shafikuzzaman Sajib et al. (2021) Situational analysis of forensic mental health in Bangladesh. PREPRINT (Version 1) available at Research Square. <https://doi.org/10.21203/rs.3.rs-1027893/v1>
28. Hayat A (2020) Moner Bondhu: A friend in need. The business standard. Available: <https://www.tbsnews.net/feature/panorama/moner-bondhu-friend-need-164950>. (Accessed: Mar. 27, 2024)
29. Kader R (2022) Anowar Hossain: The making of DocTime, the future of digital healthcare in Bangladesh, and entrepreneurship (Part II),” Future Startup. Available: <https://futurestartup.com/2022/08/02/anowar-hossain-the-making-of-doctime-the-future-of-digital-healthcare-in-bangladesh-and-entrepreneurship-part-ii/>. (Accessed: Mar. 27, 2024)
30. Das S, Namasudra S (2023) MACPABE: Multi-Authority-based CP-ABE with efficient attribute revocation for IoT-enabled healthcare infrastructure. *Int J Network Manage* 33(3):e2200
31. Sarkar S, Saha K, Namasudra S, Roy P (2015) An efficient and time saving web service based android application. *SSRG Int J Comput Sci Eng (SSRG-IJCSE)*. (8):18–21
32. Datta S, Namasudra S (2024) Blockchain-Based smart contract model for securing healthcare transactions by using consumer electronics and Mobile-Edge computing. *IEEE Trans Consum Electron* 70(1):4026–4036. <https://doi.org/10.1109/TCE.2024.3357115>

# Decentralized Energy Grid System Using IoT and Blockchain: A Sustainable Future



Saswati Debnath , Vedavati Patil, and Dhruv Agrawal

**Abstract** The global pursuit of sustainable energy solutions has led to the exploration of innovative technologies aimed at revolutionizing energy management and distribution. This paper presents integration of Blockchain and Internet of Things (IoT) technologies in decentralizing energy management with a focal point on transmitting power data from small-scale renewable power stations to a central power plant. IoT devices strategically monitor energy production from renewable energy sources, such as wind turbines, solar panels, and microgrids, enabling real-time data analysis to maximize efficiency. These transactions, which are automated by smart contracts, ensure trust. Interoperability, data security, and user interfaces are all covered as the technical architecture required for seamless integration in this paper. It is important that regulations need to be changed to make room for creative energy solutions. This framework highlights sustainability, encourages energy independence, and lessens dependency on centralized power by supporting renewable energy sources. By combining blockchain technology, IoT, and sustainable energy practices, the proposed system aims to transform the way of energy distributing, improving security, resilience, and sustainability while giving to communities more control over how much energy they use and produce.

**Keywords** Decentralized energy monitoring · Internet of things · Blockchain · Real time monitoring

---

S. Debnath  · V. Patil · D. Agrawal

Department of Computer Science and Engineering, Alliance University, Bengaluru, Karnataka, India

e-mail: [debnath.saswati123@gmail.com](mailto:debnath.saswati123@gmail.com)

V. Patil

e-mail: [pvedavatibtech21@ced.alliance.edu.in](mailto:pvedavatibtech21@ced.alliance.edu.in)

D. Agrawal

e-mail: [adhruvbtech21@ced.alliance.edu.in](mailto:adhruvbtech21@ced.alliance.edu.in)

## 1 Introduction

A major energy revolution is currently taking place in the modern world, driven by the urgent need for greater energy independence, resilience, and sustainability. As the world's population expands and need for energy increases, it is becoming increasingly apparent that traditional centralized energy systems will not be able to satisfy the demands of the future.

The global energy landscape stands at a pivotal crossroads, with the imperative for sustainability driving unprecedented shifts in energy management and distribution. Conventional centralized energy systems, reliant on finite fossil fuel resources, face mounting pressure from environmental concerns and the urgent need to mitigate climate change. In response, the exploration of innovative technologies has accelerated, with a particular emphasis on Blockchain and Internet of Things (IoT) solutions, poised to reshape the future of energy infrastructure. The background of this research is rooted in the urgent need to address the shortcomings of traditional centralized energy systems and accelerate the transition towards decentralized, renewable energy solutions. Centralized grids, characterized by large-scale power plants and extensive transmission networks, are susceptible to inefficiencies, vulnerabilities, and environmental impacts.

The main aim of this work is to increase sustainability and resilience. This research seeks to explore the potential of Blockchain and IoT technologies in decentralizing energy management. The motivation stems from the recognition of the transformative power of these technologies in overcoming longstanding challenges in energy infrastructure, including grid congestion, lack of transparency, and inefficiencies in resource allocation. By harnessing the capabilities of Blockchain and IoT, this research aims to contribute to the development of more efficient, equitable, and sustainable energy systems. While the original scope of the work encompassed the integration of Blockchain technology, challenges encountered during implementation necessitated a refined focus on IoT solutions. Thus, this research primarily explores the applications of IoT technology in decentralized energy management, including data collection, transmission, and real-time monitoring. The scope extends to a comprehensive examination of IoT methodologies, challenges, and performance metrics within the context of decentralized energy systems, with implications for scalability, reliability, and regulatory compliance. The advantage of decentralized energy grid system is to improve efficiency in energy use, encouragement of renewable energy, clear and safe transactions, the empowerment of the community, sturdiness and flexibility. The most significant contributions of this research are as follows:

- I. The IoT-based energy management system is the main aim of this work. By demonstrating the effectiveness of IoT devices in collecting, transmitting, and analyzing power generation data from renewable sources, this research contributes to the development of more efficient and sustainable energy management systems.
- II. Another significant contribution of this research is the identification and mitigation of challenges associated with blockchain integration. While originally

intended to incorporate blockchain technology, challenges encountered during implementation led to a refined focus on IoT solutions.

The structure of the paper includes literature review in Sect. 2, methodology in Sect. 3. Sections 4 and 5 provide results analysis and conclusion respectively.

## 2 Literature Review

### ***2.1 Blockchain, IoT, and AI Utilized in Smart Grids to Distribute Energy Resources (2020)***

The study presented how blockchain, IoT, and AI integrated into smart grids to manage data and electricity flow in two directions within electricity system networks (ESNs) [1]. It highlighted how Distributed Energy Resources (DER) are transforming conventional grids and how AI, IoT, and Block chain enable optimal SG operations. This research demonstrated that the technologies work together to enhance the sustainability, dependability, security, and real-time monitoring of Smart Grid services.

### ***2.2 Block Chain for Managing the Internet of Energy: Overview, Issues, and Solutions (2020)***

Miglani et al. (2020) have explored the use of blockchain technology to manage the Internet of Energy (IoE) [2]. The paper emphasized the revolutionary effect of blockchain in supporting distributed and decentralized solutions while concentrating on resolving issues with centralized IoE architecture. Through a thorough analysis of blockchain technology used to manage renewable energy, electric vehicles (EVs), and smart grid advancements, the study highlighted the improved automation, security, and privacy features. The paper has given a useful summary of how block chain enables complex energy transactions, automated data exchange, and peer-to-peer (P2P) energy trading.

### ***2.3 Evaluation of Groundbreaking Techniques and Systems Advised for Upcoming IoT Infrastructure (2021)***

This paper [3] has presented revolutionizing global connectivity of IoT. The authors have proposed enhanced interoperability, scalability, integrity, and accessibility for

IoT infrastructure. The paper explored architectures, technologies, applications, challenges, protocols, and opportunities in order to tackle the complex problem of securing IoT systems. Important factors like interoperability, scalability, security, and energy efficiency are all included in the thorough taxonomy that is provided.

## **2.4 *Following Sustainability Promises in the Energy Transition Path: Using Blockchain Technology to Boost Renewable Certificates (2022)***

This paper presented the importance of the corporate transition towards sustainability and the attainment of net-zero carbon goals through the procurement of electricity [4]. The authors have drawn attention to the structural problems that currently plague Guarantees of Origin (GOs), focusing on issues with additionality, transparency, and intricate administrative structures. The suggested remedy entails using blockchain networks to address these problems, promoting transparency, encouraging investment in renewable energy sources, and optimizing procedures to satisfy the demands of the energy sector going forward.

## **2.5 *Blockchain's Untapped Potential for Sustainable Supply Chains (2020)***

A thorough analysis of how blockchain technology might promote sustainability in a number of different industries has been presented in this paper [5]. The review examined how it has transformed supply chains, energy, tourism, healthcare, and finance. The authors have presented the complex relationship between blockchain technology and social, economic, and environmental sustainability through a methodical examination of 37 documents. The study identified crucial research gaps that entrepreneurs and researchers should fill in addition to offering useful implications that showed improvement on business profits and reputation.

## **2.6 *Decentralized Energy Systems Based on IoT (2022)***

In order to meet sustainability goals, the energy sector is shifting from traditional centralized systems to decentralized structures. The article employed keyword biometric analysis to highlight advancements in industry 4.0, IoT, and information and communication technologies [6]. This offered insightful information about the state of global research on decentralized energy systems today. This adds to our understanding of how sustainable energy transformation might be accomplished.

## ***2.7 Blockchain-Based Peer-To-Peer Energy Trading (2023)***

According to this study [7], sustainability and an analysis of the environmental impact of blockchain-based energy trading can improve the debate around this technology. By addressing data protection, regulatory frameworks, and interoperability, a comprehensive perspective is offered. Optimization methodology would be provided by real-world examples or pilot projects showcasing successful blockchain applications for peer-to-peer energy trading. Beyond the present situation, it could be beneficial to look at potential sustainability and environmental impacts of implementing peer-to-peer energy trading based on blockchain technology.

## ***2.8 Blockchain for Low-Carbon Smart Energy Systems to Trade Electricity Data (2024)***

The authors of "Blockchain for Electricity Data Trading in Low-Carbon Smart Energy Systems," offered a novel strategy for trading electricity data in low-carbon power systems that combines blockchain technology and price games [8]. The approach consisted of a thorough framework with consumers, data providers, and an information system built on blockchain. Special features like digital watermarking and blockchain for copyright protection have been introduced, along with information theory for data valuation and multiobjective optimization for pricing. The method's effectiveness is examined by the experimental results, offering power enterprises a promising way to finance the cost of investing in low- carbon smart energy systems.

## ***2.9 Establishing a Decentralized, Peer-To-Peer Energy Market that Accounts for Transaction Fees, Loss, and Fairness in the Context of an Active Distribution Network (2024)***

A fully decentralized peer-to-peer (P2P) energy market is introduced in the paper "Creating a Decentralized Peer- to-Peer Energy Market for an Active Distribution Network" by Amir Zore et al. It emphasized consumer-centric models, direct bilateral trading, and considerations for power losses and transaction fees [9]. Flexibility is increased by participating in demand response programs and incorporating regional P2P marketplaces.

## ***2.10 Innovative Applications of Blockchain Technology in the Electrical Energy Sector (2024)***

The authors examined the revolutionary effects of blockchain technology on the electrical energy industry in this paper [10]. Using a multidisciplinary methodology that included a literature review, technological analysis, case studies, and stakeholder interviews, the research highlighted blockchain's potential in addressing issues related to energy management and distribution. Aspects like energy consumption and scalability have been discussed alongside important applications like grid management, renewable energy integration, and peer-to-peer energy trading. The results showed the continued need for research and development in this dynamic field and suggested that blockchain holds promise for developing more democratic, efficient, and sustainable energy systems despite obstacles.

## ***2.11 Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications***

This paper has presented the IoT with an emphasis on its enabling technologies, protocols, and applications. The authors have given a thorough understanding of IoT development, problems, and possible solutions [11]. They talk about developments in communication technologies, and smart sensors, emphasizing the move toward direct cooperation amongst smart sensors. The survey helps researchers and developers by providing insights into pertinent protocols and application issues, even in the face of obstacles like interoperability and security concerns.

## ***2.12 Bit Coin's Growing Energy Problem***

The authors examined the rising energy consumption linked to Bitcoin mining. It looks into the underlying causes, such as the increasing computational complexity that mining operations require and the resulting spike in energy consumption [12]. The study emphasized how Bitcoin's energy usage has an impact on the environment, especially in terms of carbon emissions and sustainability. The authors clarified the extent of Bitcoin's energy consumption and its expected future course through data analysis. Additionally, they suggested possible solutions to lessen the environmental impact of Bitcoin, like switching to energy-efficient mining algorithms and incorporating renewable energy sources.

## ***2.13 The Nature of Human Design Thinking in the Context of Bringing Systems to Life***

Forrester, J. et al. proposed human design thinking in relation to bringing systems to life [13]. The nature of design thinking and its use in system development are examined in this paper. It explores the nuances of design processes that are human-centered, highlighting the significance of empathy, creativity, and teamwork. The authors have highlighted the value of human viewpoints in system design to improve usability and user experience through case studies and examples. They have addressed issues pertaining to user requirements and system complexity, putting forward solutions that put human-centric design principles first.

## ***2.14 Low-Cost Solution to the Grid Reliability Problem with 100% Penetration of Intermittent Wind, Water, and Solar for All Purposes***

A method to maintain grid reliability during the switch to 100% renewable energy is described in "Low-cost solution to the grid reliability problem with 100% penetration of intermittent wind, water, and solar for all purposes" was proposed by Jacobson et al. [14]. Their method combined energy storage, grid management, and renewable energy technologies, demonstrating the viability of reliability without the use of nuclear or fossil fuels. The paper provided useful insights for policymakers and energy planners by advocating for a shift to renewable energy through the optimization of renewable resources, enhancement of grid flexibility, and implementation of efficient storage.

## ***2.15 Blockchain Challenges and Opportunities: A Survey***

A thorough review of the literature on the growing significance and interest of local energy communities in the energy sector has been discussed in the paper [15]. It has focused on a number of local energy community topics, such as their definition, traits, advantages, difficulties, and legal environments. The authors have provided valuable insights into the state of research and practice in this field today through a thorough review and analysis of the body of existing literature. The study highlighted how local energy communities can help ease the shift to a more decentralized and sustainable energy system.

## ***2.16 Blockchain-Based Smart Contract Model for Securing Healthcare Transactions by Using Consumer Electronics and Mobile Edge Computing***

Blockchain technology is becoming a popular answer to a number of issues facing the healthcare industry. Its uses in the healthcare industry include everything from protecting patient information to boosting supply chain transparency for pharmaceuticals.

In this paper [16], health data has been gathered and processed by consumer electronics devices and then uploaded to a blockchain network.

## ***2.17 Blockchain-Based Secure and Scalable Supply Chain Management System to Prevent Drug Counterfeiting***

Drug distribution to customers must be dependable and safe, which requires a well-managed Supply Chain Management (SCM) system. The difficulty of tracking a drug along the supply chain in conventional SCM systems contributes to the likelihood of drug counterfeiting. In addition, the scalability problem with traditional SCM systems derives from their restricted information processing capacity. In order to address the issues with conventional SCMs, a blockchain-based plan utilizing the InterPlanetary File System (IPFS) has been presented in this study [17].

## ***2.18 User Revocation-Enabled Access Control Model Using Identity-Based Signature in the Cloud Computing Environment***

These days, a lot of data is kept on cloud servers for cross-domain sharing. Concerns regarding data confidentiality arise while storing or sharing data on the cloud due to the growing number of security vulnerabilities. The revocation of rogue users is an unresolved yet crucial challenge in cloud data-sharing platforms. Revocation is sometimes accomplished by periodically upgrading users' private keys. With this strategy, as the user base grows, the Key Generation Center's (KGC) burden increases. In this work [18], an effective Revocable Identity-Based Signature (RIBS) system has been developed, in which an External Revocation Server (ERS) is given the revocation capabilities. This suggested plan offers restricted access control by limiting system resource access to only those users who have not had their privileges revoked.

Table 1 represents key points of the few papers listed in the literature review.

**Table 1** Key-points from literature

Paper title	Year	Key-points
Blockchain, IoT, and AI Utilized in Smart Grids to Distribute Energy Resources	2020	The paper emphasizes how Distributed Energy Resources (DER) and the Internet of Things (IoT) and blockchain are transforming traditional grids and improving grid management features like bidirectional electricity flow and real-time monitoring
Block chain for managing the Internet of Energy: Overview, issues, and solutions	2020	The paper discusses important concerns with centralized IoE architecture and demonstrates how blockchain can improve automation, security, and privacy, among other issues
Evaluation of Groundbreaking Techniques and Systems Advised for Upcoming IoT Infrastructure	2021	The taxonomy's recognition of security and energy efficiency as crucial factors demonstrates a thorough comprehension of the major problems that IoT infrastructure faces
Following Sustainability Promises in the Energy Transition Path: Using Blockchain Technology to Boost Renewable Certificates	2022	The paper is forward-thinking and pertinent to current and upcoming developments because it makes suggestions about how blockchain can be used to meet future demands in the energy sector
Blockchain's Untapped Potential for Sustainable Supply Chains	2020	Industry practitioners aiming to use blockchain for sustainability will find the findings useful as they provide practical implications for how blockchain can enhance business profits and reputation
Decentralized Energy Systems based on IoT	2022	Emphasis on information and communication technologies (ICT), which highlights their significance in enabling decentralized energy systems
Blockchain-based peer-to-peer energy trading	2023	Future energy strategies should take this into account. The paper focuses on how decentralized energy markets can help with the development of cleaner energy sources
Establishing a peer to peer energy market that accounts for transaction fees, loss, and fairness in the context of an active distribution network	2024	The proposed system's effectiveness is demonstrated through a numerical study of an IEEE 13-bus distribution network, which provides empirical evidence to bolster the theoretical framework and bolster its credibility
Innovative Applications of Blockchain Technology in the Electrical Energy Sector	2024	In addition to highlighting the ongoing need for research and development, the paper makes recommendations for future study topics and promotes continuous innovation in this rapidly evolving field

### 3 Proposed Methodology

The methodology section outlines the systematic approach to collect data from renewable power stations, transmit it to the main power plant, and integrate IoT devices into the energy management system.

### **3.1 Data Collection from Renewable Power Stations**

The data collection process involves several stages, each of which is crucial for obtaining accurate and reliable data.

**Selection of Power Stations:** A comprehensive review of available renewable power stations is conducted to identify suitable candidates for inclusion in the study. Factors such as geographic location, energy production capacity, and accessibility are considered in the selection process.

**Deployment of Monitoring Equipment:** Specialized monitoring equipment, including sensors, data loggers, and smart meters, is strategically deployed at each selected power station. These devices are carefully calibrated and configured to accurately measure key parameters such as power output, voltage, current, and environmental conditions.

**Data Acquisition Protocols:** Standardized protocols for data acquisition are established to ensure consistency and reliability across all monitoring devices. This includes defining sampling frequencies, data transmission intervals, and data format specifications to facilitate seamless integration and analysis.

**Quality Assurance Measures:** Rigorous quality assurance measures are implemented to validate the accuracy and integrity of the collected data. Regular calibration checks, sensor recalibration, and data validation procedures are performed to identify and rectify any anomalies or discrepancies.

### **3.2 Data Transmission to Main Power Plant**

The transmission of data from renewable power stations to the main power plant involves the following steps:

**Communication Infrastructure:** A robust communication infrastructure is established to facilitate the transmission of data over long distances. This may include the deployment of wireless communication technologies such as Wi-Fi, cellular, or satellite communication, depending on the geographic location and accessibility of the power stations.

**Data Transmission Protocols:** Standardized data transmission protocols are employed to ensure efficient, secure, and reliable communication between remote power stations and the main power plant. Protocols such as TCP/IP, MQTT, or LoRaWAN are selected based on their compatibility, scalability, and security features.

**Data Aggregation and Synchronization:** Data collected from multiple power stations are aggregated, synchronized, and harmonized to create a unified dataset for analysis. Time synchronization protocols such as Network Time Protocol (NTP) or Precision Time Protocol (PTP) are utilized to ensure temporal alignment of data from disparate sources.

**Real-time Monitoring and Control:** The transmitted data is processed in real-time to enable continuous monitoring and control of power generation assets.

Advanced analytics algorithms, machine learning models, and predictive analytics techniques are employed to identify patterns, anomalies, and opportunities for optimization.

### ***3.3 Integration of IoT Devices***

The integration of IoT devices into the energy management system is a multifaceted process that involves the following components:

**Selection of IoT Devices:** A comprehensive evaluation of available IoT devices is conducted to identify those that best meet the requirements of the energy management system. Considerations include compatibility with existing infrastructure, scalability, reliability, and cost-effectiveness.

**Device Deployment and Configuration:** Selected IoT devices are deployed at strategic locations within the energy infrastructure and configured to interface seamlessly with other components of the system. This involves programming device firmware, configuring network settings, and integrating with data acquisition systems.

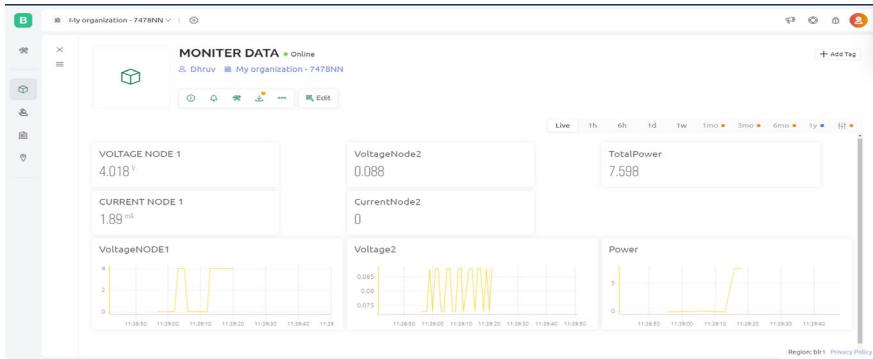
**Data Processing and Analysis:** Data collected from IoT devices are processed, analyzed, and visualized to extract actionable insights and inform decision-making processes. Advanced data analytics techniques, including time series analysis, anomaly detection, and pattern recognition, are employed to identify trends, optimize performance, and enhance efficiency.

**System Integration and Interoperability:** IoT devices are integrated into the existing energy management system to enable interoperability and seamless data exchange. Open standards such as Message Queuing Telemetry Transport (MQTT), Constrained Application Protocol (CoAP) are leveraged to facilitate communication between devices, systems, and applications. Middleware platforms and gateways may be deployed to bridge communication gaps and facilitate data exchange between heterogeneous devices and protocols.

## **4 Results and Discussions**

### ***4.1 Analysis of Data Transmission Efficiency***

The analysis of data transmission efficiency evaluates the effectiveness of transmitting power generation data from renewable power stations to the main power plant. Key metrics for evaluation include data transmission latency, reliability, and bandwidth utilization. Results indicate that the implemented communication infrastructure and protocols facilitate timely and reliable transmission of data, enabling



**Fig. 1** Analysis of data transmission efficiency

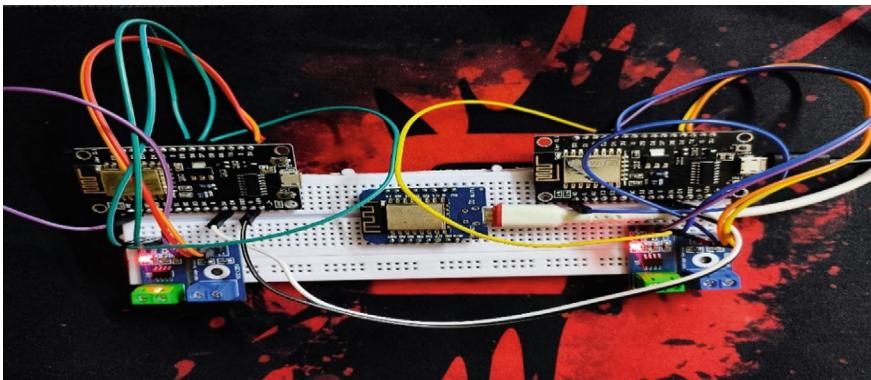
real-time monitoring and control of power generation assets. Figure 1 represents the analysis of data transmission efficiency.

## 4.2 Evaluation of IoT Device Performance

The evaluation of IoT device performance assesses the reliability, accuracy, and responsiveness of deployed IoT devices in monitoring and controlling energy assets. Results demonstrate that IoT devices effectively collect and transmit data from renewable power stations, providing granular insights into power generation metrics and environmental conditions. However, challenges such as limited battery life and connectivity issues may impact device performance in remote or harsh environments.

## 4.3 Comparison with Blockchain-Integrated Systems

A comparison with blockchain-integrated systems provides valuable insights into the advantages and limitations of both approaches. While blockchain technology offers benefits such as enhanced security, transparency, and decentralization, challenges such as scalability and regulatory compliance hinder widespread adoption. In contrast, IoT-based solutions offer real-time monitoring, control, and optimization capabilities without the complexities associated with blockchain integration. The results suggest that IoT-based energy management solutions may offer a pragmatic alternative to blockchain-integrated systems, particularly in environments where scalability and regulatory constraints are significant concerns. Figure 2 represents the device deployment and configuration.



**Fig. 2** Device deployment and configuration

## 5 Conclusion

This research has demonstrated the feasibility and effectiveness of IoT-based energy management solutions in decentralizing energy systems. By collecting and transmitting power generation data from renewable sources to a central power plant, IoT devices enable real-time monitoring, control, and optimization of energy assets. Despite challenges encountered in blockchain integration, the findings underscore the potential of IoT technologies to address key challenges in energy management and facilitate the transition towards more sustainable energy practices. The implications of this research for sustainable energy practices are significant. By enabling decentralized energy management, IoT technologies offer opportunities to increase energy efficiency, reduce carbon emissions, and enhance grid resilience. The findings suggest that pragmatic approaches focusing on IoT solutions can overcome barriers to adoption and accelerate the deployment of decentralized energy systems.

## References

1. Shaurath Chopra, Maria Malvoni, Nallapanneni Manoj Kumar (2020) Blockchain, IoT, and AI utilized in smart grids to distribute energy resources, *Energies* 2020, 13(21), 5739
2. Arzoo Miglani, Neeraj Kumar, Vinay Chamola, Sheralli Zeadally (2020) Blockchain for internet of energy management: review, solutions, and challenges. *Computer Communications—Elsevier*, 151, 395–418
3. Arun Kumar, Jose Manual-Brenosa, Sharadh Sharma (2021) Evaluation of groundbreaking techniques and systems advised for upcoming IoT infrastructure, *Sustainability* 2022, 14(1), 71, 5739
4. Orestis Delardas, Panagiotis Giannos (2022) Following sustainability promises in the energy transition path: using blockchain technology to boost renewable certificates. Orestis Delardas and Panagiotis Giannos, *Sustainability* 2023, 15(1), 258

5. Vincenzo Varriale, Antonello Cammarano, Francesca Michelino (2020) Blockchain's untapped potential for sustainable supply Chains, *Sustainability* 2020, 12(22)
6. Marta Bieganska (2022) Decentralized Energy System based on IoT. *Energies* 2022, 15(21), 7830
7. Whig Pawan (2023) Blockchain-based peer-to-peer energy trading— Elsevier, 2023, 171–188
8. Bonan, Yushuai Li (2024) Blockchain for Low-Carbon smart energy systems to trade electricity data. *IEEE*, 09 Huang 1–11
9. Amir Zore (2024) Mehdi mehadinejad and Mehrdad Abedi: Establishing a decentralized, peer-to-peer energy market that accounts for transaction fees, loss, and fairness in the context of an active distribution network. *Applied Energy*, 2024—Elsevier, 122527
10. Rakhmonov IU, Kurbonov NN (2024) Innovation application of blockchain technology in the electrical energy sector. *Mod Sci Res* 3(1):1–5
11. Al-Fuqaha A, Guizani M, Mohammadi M, Aledhari M, Ayyash M, Internet of things: a survey on enabling technologies, protocols, and applications. *IEEE Commun Surv & Tutor*, 17(4), 2347–2376
12. De Vries A, Timmer J (2019) Bit coin's growing energy problem: Joule, 2:5, 16 May 2018, Pages 801–805
13. Forrester J, Waterworth EL, Waterworth JA (2017) The Nature of human design thinking in the context of bringing Systems to life. *Int J Des* 11(1):11–29
14. Jacobson MZ, Delucchi MA, Cameron MA, Mathiesen BV (2015) Low-cost solution to the grid reliability problem with 100% penetration of intermittent wind, water, and solar for all purposes. *Proceeding of the national academy of sciences* 112(49):15060–15065
15. Schorlemmer H, Gidlund M, Luque A, Nordström L (2020) Blockchain challenges and opportunities: A survey *International Journal of web and grid services Local Energy Communities: A Literature Survey*. *Energies* 13(14):3518
16. Tarun Kumar,Prabhat Kumar,Suyel Namasudra,User Revocation-Enabled Access Control Model Using Identity-Based Signature in the Cloud Computing Environment, *International Journal of Interactive Multimedia and Artificial Intelligence*,<https://www.ijimai.org/journal/bibcite/reference/3451>
17. Datta S, Namasudra S (2024) Blockchain-based secure and scalable supply chain management system to prevent drug counterfeiting. *Cluster Computer*. <https://doi.org/10.1007/s10586-024-04417-3>
18. Datta, Sagnik & Namasudra, Suyel. (2024). Blockchain-Based Smart Contract Model for Securing Healthcare Transactions by Using Consumer Electronics and Mobile Edge Computing. *IEEE Transactions on Consumer Electronics*. PP. 1–1. <https://doi.org/10.1109/TCE.2024.3357115>

# Detection and Analysis of Cyber-Attacks on IoT Network Devices



Bashir Zak Adamu , Ilhan Fırat Kilincer , and Fatih Ertam

**Abstract** One of the most pressing concerns in network forensics is the detection of cyber-attacks in the Internet of Things (IoT) networks and their devices. Traditional intrusion detection systems based on signature rules are unable to detect current attack types. Hence, the need to urgently develop advanced methods for classifying IoT network traffic that can swiftly detect cyber-attacks becomes inevitable. This research aims to develop machine learning algorithms for cyber-attack detection in IoT-based networks, by analyzing the traffic data composed from the network itself. An ideal IoT network was implemented solely for the attack scenarios and generation of a dataset. For this study, both the structure and security of IoT networks were investigated in detail, by utilizing an IoT network created in real environment. Attack scenarios have been created for IoT devices in the real environment created. An IoT security data set was created by collecting the network flows obtained as a result of the attacks. On the created data set, classification results close to 100% accuracy values were obtained with machine learning algorithms. The data set obtained in the study has been published publicly so that other researchers can use it.

**Keywords** IDS · Cyber security · IoT · Machine learning

## 1 Introduction

For more than a century, mankind has been using automatic sensors and controls, calling them sensors (from the word feel and measure), and has been using them to connect computers. Furthermore, we are currently amidst a remarkable technological revolution that offers unprecedented opportunities but also poses the huge risks associated with these changes. It is now possible to use extremely inexpensive controllers and sensory devices with amazingly exponential capability, availability, and growing varieties. Following their unconditional usefulness and economic feasibility, technologies based on IoT in green buildings, environmental monitoring, health and

---

B. Z. Adamu · I. F. Kilincer · F. Ertam ()

Department of Digital Forensics Engineering, Firat University, Elazig, Türkiye  
e-mail: [fatih.ertam@firat.edu.tr](mailto:fatih.ertam@firat.edu.tr)

safety monitoring of facilities have already been deployed. We buy cars with already built-in sensors and controls or, if you prefer, with the Internet of Things (although cars, for the most part, are not directly linked to the web). With the continuous unprecedented widespread of IoT technology, they became a soft target for cyber-attacks due to their nature, mode of operation, and availability. As such, cyber-attacks on IoT networks and devices became prevalent everywhere, making it a difficult concern for forensic experts. Hence, the best way for tackling these attacks is to detect them in real-time while they occur, a terminology known as “live forensics” [1].

According to [2, 3] an IDS in IoT is a tool that monitors IoT networks by observing data in the network traffic for identification and protection against intrusions that can threaten the security of information systems within the network architecture. Such systems are used by practitioners for the discovery of security vulnerabilities as well as anomalous/malicious activities. Most of these IDS systems accomplish their task by using passive traffic collection and analysis. Since the presence of IDSs does not invalidate the use of primary security components that serve as the first line of network defense, IDSs are tagged by security practitioners as a second line of network defense [4].

Intrusion Detection Systems (IDSs) are one promising component of the live forensic approach for monitoring IoT networks and are essentially effective at the cyber level. IDSs deployed in IoT environments analyze network traffic packets and produce real-time responses [5]. But for these IDSs to be effective, they must operate under rigorous IoT conditions of low energy, low process capacity, quick response, and eminently gigantic volumes of information handling. Hence, augmenting IoT embedded IDSs significantly requires a continuous and serious understanding of the security vulnerabilities posed by IoT networks.

## 1.1 Problem Definition

Firstly, according to 2019 data from PwC (an international network of companies offering consulting and audit services), the IoT is currently an actively developing technology amassing trust for changing the business models of both companies and entire industries [6]. In the presented rating of Forbes [7], and the publication [8], “Internet of Things” is ahead of technologies such as artificial intelligence, robotics, augmented and virtual reality, with a projected number of connected device base of approximately 42.62 billion by 2022 which is expected to increase to 76 billion by 2030. According to [9], the machine-to-machine (M2M) connections count is expected to reach 27 billion by 2024.

Secondly, with the hike in IoT devices globally, all objects that are part of the IoT become a target and are under threat: objects of the Department of Defense and critical infrastructures such as employees, equipment, and other properties may be used to initiate IoT security attacks. If not looked into appropriately, adversaries who are determined may end up undermining the security and processes that will bring

down systems like power grids, medical access, and communication. One way or another, all these reputable departments are closely studying the applications of IoT in various systems. If no action is taken to get around the hike in IoT device base, the problem will grow exponentially further, and the growth of IoT attacks could sweep these structures like a tidal wave.

Furthermore, IoT attack detection has recently proved to be a pressing issue in the area of securing corporate networks. Large-scale outbreaks of network worms, DDoS attacks from botnets, automated tools for finding vulnerabilities in networks—all this makes securing IoT networks a very laborious task. Nowadays it is difficult to find a network that lacks such active means of preventing attacks as antivirus, firewall, and intrusion prevention systems. Unfortunately, active countermeasures alone are not enough. Therefore, in addition to them, passive means of fighting attacks are used as network intrusion detection systems.

## ***1.2 Motivation***

Connecting every smart object to each other as well as the internet results in numerous security and privacy challenges. Confidentiality, Authenticity, and Data Integrity as basic security elements are threatened because most IoT devices lack security by design. A malicious adversary may not even need to hack most components of the IoT because they lack authentication components. Hence, the need to detect and prevent such attacks in real-time becomes paramount.

Furthermore, Intrusion detection tools that use only the signature-based intrusion detection method in their operation cannot detect new types of attacks or modifications of old ones; therefore, the task of developing better algorithms for detecting IoT network traffic anomalies become paramount. Also, there are researches that suggest the application of machine learning principles to develop up-to-the-task IDSs that can eliminate the difficulties faced by the traditional approaches.

## ***1.3 Contribution***

The advantages/contributions of the proposed method are as follows;

- The design and implementation of a real-time IoT network.
- Practical execution of various IoT cyber-attacks on the IoT network devices.
- Self-generation of dataset used for classification.
- Design and implementation of an automated Machine learning algorithm for detecting and classifying the experimental attacks for security and forensic purposes.

- The dataset used in the study is publicly shared for contribution to cyber security works in the future.<sup>1</sup>

## 1.4 Paper Structure

The paper is structured as follows; Sect. 1 as seen above is the Introduction, where the concept, problem statement, motivation, and contributions were presented; Sect. 2 buttresses the validity of the research by portraying related works from different authors; Sect. 3 is the Method which encompasses the main work, the proposed framework and its implementation is explained here including the IoT network architecture, executed attack categories, dataset compilation, preprocessing, and training/testing.; Sect. 4 entails the experimental results and discussion, where comparison of the said results were also conducted; Sect. 5 which is the last section concludes the research by providing conclusive remarks as well as recommendations for future works.

## 2 Related Works

Most IoT security researchers see machine learning techniques as part of their approach to detecting attacks and anomalous device behavior. The main advantages of machine learning over traditional signature techniques are high performance and scalability for growing amounts of data, and the ability to automatically select useful features from raw data.

The work [10] proposes the concept of a device, which is a small module that can be embedded both in an extensive IoT network and in a network with limited resources, thereby becoming a monitoring system and allowing to organize an uninterrupted check of the Internet of Things network for attempts to implement an attack on a given network. This module consists of four elements responsible for capturing traffic inside the IoT network, filtering it, and issuing a verdict on an attempt to implement one or another attack on the devices of this network.

Similarly, the paper [11] explores the potential of recurrent neural networks with LSTMs for detecting IoT malware. The developed model is compared to classifiers based on traditional machine learning techniques: Support Vector Machine (SVM), Naive Bayes Classifier, Random Forest, Adaptive Boosting (AdaBoost), K-Nearest Neighbors (KNN). Analysis shows that the deep learning approach gives the best possible results. However, no comparison with other deep learning models was made.

In a study [12], the authors propose a unique anomaly detection system for industrial IoT systems that uses an autoencoder and a deep feedforward neural network. The developed model is compared to the properties of several developed anomaly

---

<sup>1</sup> [https://github.com/fatihtam/IoT\\_dataset](https://github.com/fatihtam/IoT_dataset).

detection methods such as Deep Trust Web, Recurrent Network, DNN, and EnsembleDNN. At the same time, the presented model was evaluated on different hardware and software with different subsets of initial data.

The article [13] proposed a decentralized cloud-based deep learning environment for discovery prevention of phishing and botnet attacks on smart devices. The developed RNN-LSTM model is compared to the deep learning model developed by other researchers. The main drawback of comparative analysis in this task is that the DNN model under consideration is not evaluated on the same dataset.

The authors in [14] analyzed several deep learning methods for detecting DDoS attacks. Multilayer perceptron, convolutional neural network, RNN-LSTM, CNN + LSTM-ensemble. These are further compared to machine learning techniques. Support vector machine, Bayes classifier, and random forest were considered. The authors concluded that deep learning techniques, especially recurrent networks, are more effective.

In the study by [15], methods were identified to prevent attacks on IoT networks by using an IoT monitoring system. At the same time, the study considered a special module for monitoring traffic within the Internet of Things network, which allows detecting attempts to carry out attacks on devices of this network. A machine learning approach was also employed.

In the paper [16], a novel lightweight privacy-preserving user and device authentication scheme based on cryptographic operations was presented. It prevents unauthorized users from accessing the healthcare system, maintains anonymity, resists cryptographic attacks, and maintains forward and backward secrecy. It was validated using security and performance analysis.

Still on IoT-enabled healthcare systems, the paper [17] proposed a framework to protect communication between an approved device and a gateway and keep unauthorized devices out of medical systems utilizing a mutual authentication technique that protects privacy. This method aims to provide simple and efficient network device authentication. Subsequently, the XOR, concatenation, and hash operations used in this suggested authentication method are lightweight cryptographic primitives.

The research in [18] proposed a lightweight and privacy-preserving mutual authentication scheme for wearable devices assisted by a cloud server. It can be divided into three phases: initialization, pairing and authentication. The scheme must resist forgery attack, de-synchronization attack, and so on. Three cases were included in the research, and the findings back up prior research.

Yuan-ko Huang in [19] described design of a smart cabin lighting system based on Internet of Things that aims to reduce development costs and save power consumption. This paper aims to design an IoT-based smart ship cabin lighting system based on the Internet of Things architecture. Location sensors in the ship cabin automatically identify the whereabouts of crew members. The data is then sensed by smart lighting devices and sent back to the server.

In the research [20] by Suyel Namasudra, a novel cryptosystem for a cloud-based IoT architecture based on DNA steganography and cryptography was proposed. Prior to being stored on the cloud server, the suggested cryptosystem not only conceals the data but also encrypts the private information. The experimental results demonstrate

that the proposed cryptosystem supports less distortion and the quality of the cover image is high after hiding the data in it. Furthermore, the time for data encryption, data decryption, key generation, and key retrieval of the proposed cryptosystem is less compared to the state-of-the-art schemes.

### 3 Method

With respect to the review carried out on different approaches to intrusion detection in IoT networks as seen in this paper, we decided to employ the anomaly-based intrusion detection technique and machine learning methods for this research methodology. Furthermore, to evaluate the performance of an IoT anomaly IDS, raw data is highly essential. The data should encompass various network features like flow, behavior, payload, source, and destination port number.

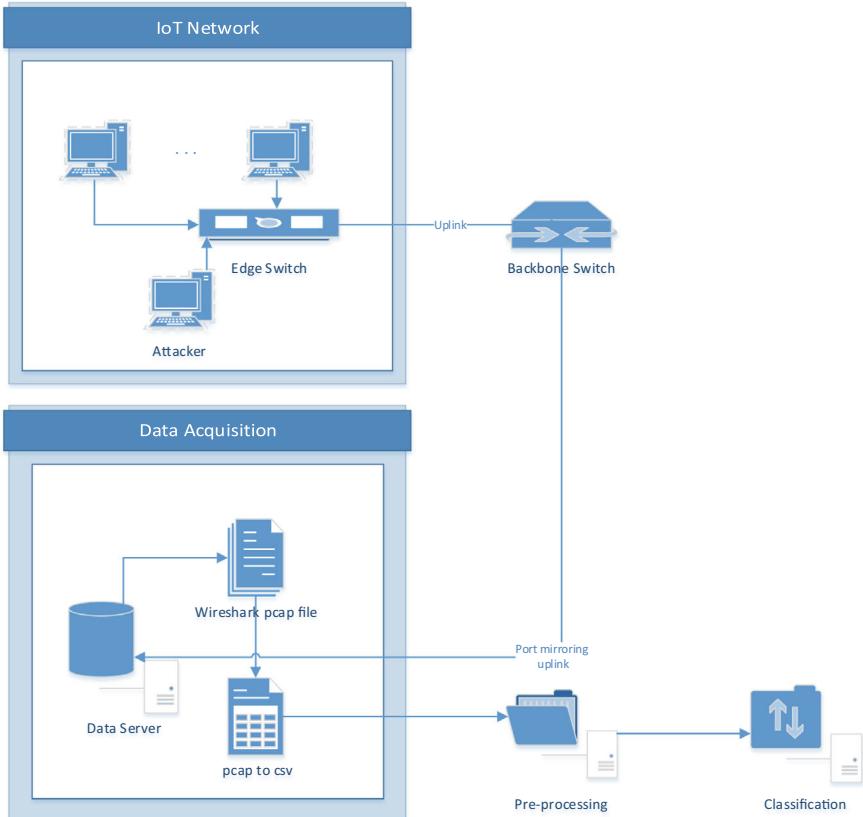
The cyber-attack categories chosen were launched on different nodes within the network and the packets were sniffed via Wireshark to generate the pcap files. The overall process starting from the attack launching stage to the final stage of prediction can be seen in the process flow Fig. 1. For further understanding of the whole methodology, the network structure, selected attacks, dataset generation, preprocessing, training, and prediction are explained below.

#### 3.1 Network Structure

The virtual machines were deployed on the Fujitsu victim PC with the required OS installed on each machine. The SBCs were also connected to the switch and configured. The default raspberryPi OS was installed on the raspberryPi SBC and the Arduino was connected to the PC via USB with the Arduino IDE application as the mediator. The Arduino compatible Ethernet board (WIZnet Ethernet WS100) was then connected to the Arduino and configured via the IDE as well. From this point, the Arduino was assigned an IP and server address for accessibility via browser. Sensors like Ultrasonic, Proximity, Buzzer, Wireless, Optical, Temperature and cameras were then connected to the Arduino as well as the RaspberryPi. The Odroid was also connected to the network just like the Arduino and some sensors also connected to it.

#### 3.2 Executed Attack Categories

Cyber-attacks were the central issue that brought about this research in the first place. The whole research is focused on finding reliable ways of detecting cyber-attacks while they happen so as to be able to prevent such attacks instantly or in the near



**Fig. 1** Proposed method

future. Owing to the fact that this research suggested the use of self-generated dataset, some cyber-attacks were selected for the research at hand. Eight attack categories were selected and launched successfully on the IoT network and its devices. The attacks were Brute force RDP, Brute force SSH, Brute force Telnet, DoS ICMP Flood, DoS SYN Flood, DoS Http get Flood, Fuzzing, and SYN Scan. An additional class of normal traffic is also included making the overall class total to be nine, as shown in the Table 1.

### 3.3 Dataset Acquisition

As explained earlier in this chapter, raw data is the primary ingredient for forensic evaluation of any IoT anomaly IDS. As such, the acquisition of the dataset for our experimentation and research became paramount since it has been proposed to be

**Table 1** Executed attack description

Attack Class	Description
Dos_Http_get_flood	Exploits seemingly-legitimate HTTP GET or POST requests
Fuzzing	Randomly feeds invalid and unexpected data inputs to stress the network
SYN_Scan	Determines the status of communication ports without full connection
Brute_Force_SSH	Brute force guessing via SSH ports
Brute_Force_RDP	Brute force guessing via RDP ports
Brute_Force_Telnet	Brute force guessing via Telnet protocol lines
DoS_ICMP_FLOOD	Overwhelms the target with ICMP ping requests
DoS_SYN_FLOOD	Rapidly sends a series of synchronized messages to the target
Normal	Normal network traffic without any compromise

self-generated. As part of the contribution for this thesis, the dataset used was self-generated from the implemented IoT network.

Before launching any cyber-attack, the normal network traffic of the connected IoT devices and the overall network traffic was sniffed via wireshark and the pcap files saved were labelled as normal. Considering the fact that eight attacks were required, the network traffic containing those attacks were also sniffed concurrently whilst the attacks were successfully executed on the go. The pcap files saved after successful execution of every attack category were also labelled according to the name of the attack. When the pcap files for all the attacks required were saved, they were all parsed using the CICFlowmeter software and the csv file for the dataset was generated.

### 3.4 Preprocessing

Data preprocessing here involves transforming the input dataset: which was made up of 80 network features (classes) into an input compliance format for algorithms to be analyzed. To represent the categorical variables in binary vector formats, the “one-hot encoding” was applied to nominal features like IP address and protocol name. To avoid instability and other unforeseen circumstances, data normalization was also passed out on the remaining features due to the imbalance between them. The values were normalized ranging from 0 to 1. Furthermore, 75% of the original dataset (6750 records) were selected as training data for all the models, while the remaining 25% (2250 records) were used for testing. At this investigative phase, high imbalance between normal and abnormal network class were encountered, which happens to be a common experience when handling anomalous traffic flow. Both the training and test data were homogenous with abnormal-to-normal ratio of 1:4. The equation for the normalization can be seen in Eq. 1.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

### 3.5 Classification and Performance Matrices

The machine learning methods utilized were also selected based on the systematic analysis of existing approaches used to identify network anomalies in cyber-security. The analysis of indicators for detecting anomalies in network traffic was further done on models from different classes as well as models from the same class that have varying number of neurons and layers. This allowed us to experimentally determine the relationship between the structure of the models and their performance. Based on the above, the following models of machine learning were analyzed: DT, SVM, KNN, and Ensemble algorithms.

The indication metrics for measuring the efficiency of detection algorithms used were: level of accuracy, prediction speed, and the training time of all algorithms. For the classification, multi-class classification was used instead of the binary classification in order to ascertain the attack category when the anomaly is detected. The multi-class classes summed up to nine classes (8 attacks + Normal). For each classification algorithm, 75% of the random records of the dataset converted to csv via CICflowmeter were used for training, while the remaining 25% were used for the accuracy assessment using a 5-block cross-validation.

## 4 Experimental Results and Discussion

The dataset (csv file) generated from the CICflowmeter initially contained 9002 records and 80 features, the 9002 records represent the attack traffic as well as the normal traffic with a total size of 11.7MB. The file was later preprocessed by deleting unnecessary features that might slow the classification process or even compromise it. The deleted features were mostly deleted because they are strings (not numbers) or have zero values all through. The deleted features can be seen in Table 2. All values were later normalized using the minimum to maximum normalization between 1 and 0 for faster training and classification by the models. After the normalization, the dataset file includes 74 features and 9000 records, where each class has 1000 records. The last feature (class) contains the labelled name of each attack traffic, hence the classes to be used for the multi-class classification algorithms.

The dataset was then successfully imported into the MATLAB environment without any error and all features and structures were intact. At this stage, the proposed classification models were applied to the dataset, after the training and testing dataset were separated. The training went smooth and good results were achieved as anticipated from the selected machine learning algorithms.

**Table 2** Altered features

Flow ID	Flow Identifier
Bwd PSH Flags	Number of times the PSH flag was set in packets travelling in the backward direction
Fwd URG Flags	Number of times the PSH flag was set in packets travelling in the backward direction
Fwd Bytes/Bulk Avg	Number of times the PSH flag was set in packets travelling in the backward direction
Fwd Packet/Bulk Avg	Number of times the PSH flag was set in packets travelling in the backward direction
Fwd Bulk Rate Avg	Number of times the PSH flag was set in packets travelling in the backward direction
Subflow Bwd Packets	Number of times the PSH flag was set in packets travelling in the backward direction

The training and testing were further reinitiated twice to see if the results will differ variably. The results of the re-run came out almost similar with differences of less than 1%.

The formulae for calculating each performance metric is given in the Eqs. 2–5.

$$\text{Accuracy} = \frac{TP + TN}{(TP + TN) + (FP + FN)} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{GeometricMean} = \sqrt{\frac{TP * TN}{(TP + FN) * (TN + FP)}} \quad (4)$$

$$F - \text{Measure} = \frac{2TP}{2TP + FP + FN} \quad (5)$$

Here, the average of the results for each algorithm was taken as seen in Table 3.

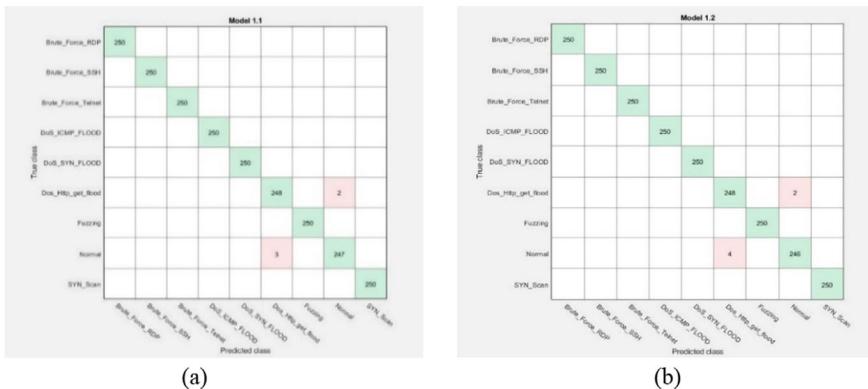
The accuracy values obtained for the classifiers used are shown in the confusion matrix in Figs. 2, 3, 4 and 5.

For the Tree algorithms, the fine tree yielded more accuracy (99.8%, Fig. 2.a) followed by the medium tree (99.7%, Fig. 2b) as can be seen in Fig. 2, and finally, the coarse tree which produced a very poor result (55.6%). For training time, the fine tree took longer (1.4451 s) and the medium tree took an average time (0.97806 s) while the coarse tree took the lesser time (0.80334).

From Fig. 3, we can see that the cubic (Fig. 3a) and quadratic (Fig. 3b) SVMs portrayed the highest accuracies of 98.5%, followed by the linear SVM (98.4%). The fine Gaussian yielded an accuracy of 95.8%, the medium Gaussian 98.6%, and finally the coarse Gaussian 96.8%.

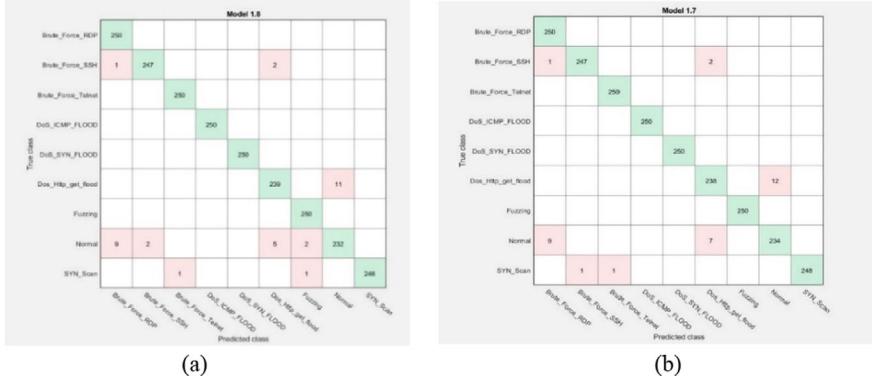
**Table 3** Average results for each algorithm

Method		Accuracy (%)	Prediction Speed obs/sec	Training Time (Sec)
	Fine Tree	99.8	110,000	1.4451
DT	Medium Tree	99.7	130,000	0.97806
	Coarse Tree	55.6	160,000	0.80334
	Linear SVM	98.4	11,000	6.0446
SVM	Quadratic SVM	98.5	7800	8.8809
	Cubic SVM	98.5	10,000	207.22
	Fine KNN	97.3	1400	3.937
KNN	Medium KNN	96.2	1300	4.3359
	Cosine KNN	96.8	1200	7.8711
	Boosted Trees	99.9	21,000	19.009
Ensemble	Bagged Trees	99.9	18,000	11.013
	Subspace KNN	99.3	120	48.567

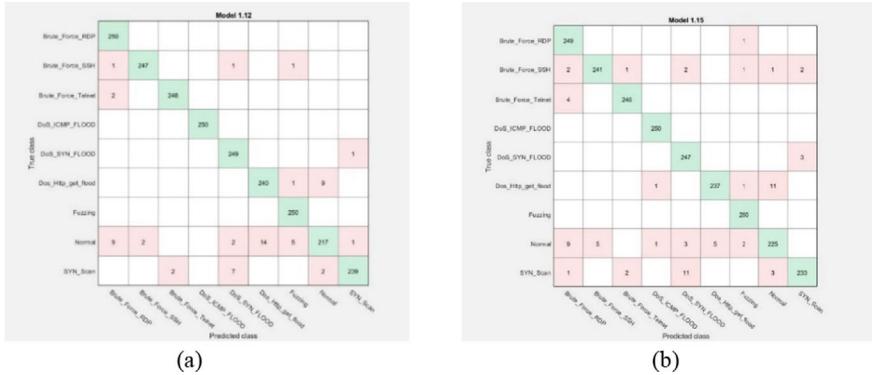
**Fig. 2** Best results of the Tree Algorithms. **a** fine tree algorithm **b** medium tree algorithm

Out of the KNN models, the fine KNN yielded the best result (97.3%, Fig. 4.a), followed by the weighted KNN (97.2%, Fig. 4.b), this can be clearly seen pictorially in Fig. 4. The medium KNN also yielded a good result (96.2%) while the coarse KNN is 89.2%, the cosine KNN finished at the accuracy of 96.8%, and finally the cubic KNN (95.8%).

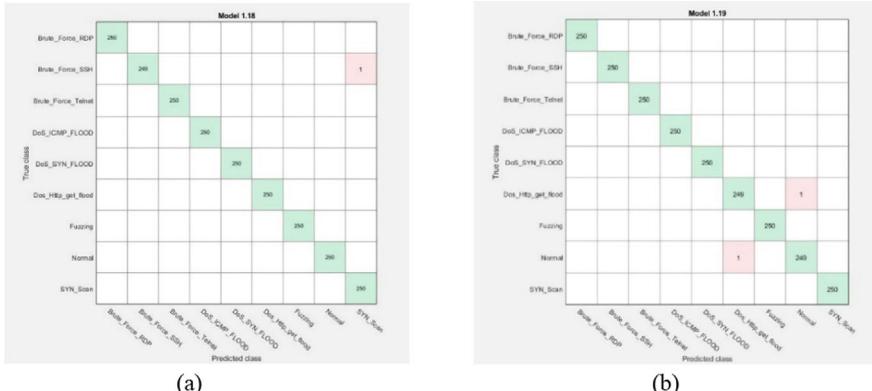
All the ensemble models produced higher accuracy rates when compared to the other models mentioned above. The boosted and bagged trees all yielded an accuracy of 99.9% each followed by the Random Under-Sampling Boost (RUSBoost) trees (99.7%) while the subspace discriminant and subspace KNN yielded 96.0% and 99.3% respectively. The success rate of this algorithm over the other algorithms is clearly visible in the boosted trees represented in Fig. 5a which mispredicted one instance only and the bagged trees mispredicted two instances only (Fig. 5b).



**Fig. 3** Best results of the SVM algorithms. **a** Cubic SVM algorithm **(b)** Quadratic SVM algorithm



**Fig. 4** Best results of the KNN. **a** fine KNN algorithm **(b)** weighted KNN algorithm



**Fig. 5** Best results of the Ensemble. **a** boosted tree algorithm **(b)** bagged tree algorithm

**Table 4** Results comparison

Method		Accuracy	Precision	Geometric Mean	F-Measure
	Fine Tree	0.9982	0.9975	0.9975	0.9976
DT	Medium Tree	0.9978	0.9975	0.7983	0.9976
	Coarse Tree	0.5564	0.5562	0.5542	0.5562
	Linear SVM	0.9843	0.9764	0.9623	0.9714
SVM	Quadratic SVM	0.9851	0.9802	0.9782	0.9842
	Cubic SVM	0.9852	0.9765	0.9736	0.9758
	Fine KNN	0.9734	0.9544	0.9476	0.9553
KNN	Medium KNN	0.9621	0.9453	0.9365	0.9437
	Cosine KNN	0.9683	0.9568	0.9384	0.9586
	Boosted Trees	0.9998	0.9995	0.9996	0.9996
Ensemble	Bagged Trees	0.9998	0.9994	0.9994	0.9994
	Subspace KNN	0.9934	0.9932	0.9931	0.9931

The results of the experiments performed suggest that most of the machine learning algorithms used have high level of accuracy in detecting large heterogeneous traffic anomalies for real-world applications. This was further confirmed by comparing the results we obtained with the results obtained from similar experimentations by other authors in various research publications.

The results obtained were further analyzed and compared in Table 4. according to their accuracies, precision, f-measure, and geometric progression.

## 5 Conclusions and Future Works

In summary, due to a large amount of metadata to be analyzed, today's intrusion detection and prevention methods require modification using various machine learning approaches to detect malicious activities in IoT network traffic or devices. A reference solution designed for the forensic purpose of detecting and preventing cyber-attacks should analyze the incoming data set of any size, quality, and depth, using artificial intelligence methods to classify traffic as legitimate or as malicious. In the latter case, the software must provide a list of threats, with their classification, probabilities, and attack trajectories to support cyber-forensic investigations.

Staying at least one step ahead of cyber-attack perpetrators is the only way to win in the war against cyber-attacks. Researchers and practitioners in the field of forensics, cyber security, and IoT security must continue to update their tools of trade. As such, the journey to attaining maximum security should be a never ending one because deploying more IoT technologies increases the number of threats since the adversaries will have more targets especially when those targets have low processing power and low-level security interfaces.

**Acknowledgements** This work has been partially supported by The Scientific and Technological Research Institution of Turkey (Türkiye Bilimsel ve Teknolojik Araştırma Kurumu, TÜBİTAK) Grant No: 123E706.

## References

1. Kilincer IF, Ertam F, Sengur A (2021) Machine learning methods for cyber security intrusion detection: Datasets and comparative study. *Comput Networks*. <https://doi.org/10.1016/j.comnet.2021.107840>
2. Elrawy MF, Awad AI, Hamed HFA (2018) Intrusion detection systems for IoT-based smart environments: a survey. *J Cloud Comput* 7:1–20. <https://doi.org/10.1186/s13677-018-0123-6>
3. Das S, Ashrafuzzaman M, Sheldon FT, Shiva S (2020) Network intrusion detection using natural language processing and ensemble machine learning. In: 2020 IEEE Symp. Ser. Comput. Intell. SSCI 2020. 829–835. <https://doi.org/10.1109/SSCI47803.2020.9308268>
4. Albulayhi K, Smadi AA, Sheldon FT, Abercrombie RK (2021) IoT intrusion detection taxonomy, reference architecture, and analyses. *Sensors*. 21. <https://doi.org/10.3390/s21196432>
5. Tahaei H, Afifi F, Asemi A, Zaki F, Anuar NB (2020) The rise of traffic classification in IoT networks: A survey. *J Netw Comput Appl* 154. <https://doi.org/10.1016/j.jnca.2020.102538>
6. PWC: 2019 IoT Survey: Speed operations, strengthen relationships and drive what's next
7. Forbes: Roundup Of Internet Of Things Forecasts
8. Shafique K, Khawaja BA, Sabir F, Qazi S, Mustaqim M (2020) Internet of things (IoT) for next-generation smart systems: A review of current challenges, future trends and prospects for emerging 5G-IoT Scenarios. *IEEE Access*. 8:23022–23040. <https://doi.org/10.1109/ACCESS.2020.2970118>
9. Kumar S, Tiwari P, Zymbler M (2019) Internet of Things is a revolutionary approach for future technology enhancement: a review. *J Big Data* 6. <https://doi.org/10.1186/s40537-019-0268-2>.
10. Casola V, De Benedictis A, Riccio A, Rivera D, Mallouli W, de Oca E (2019) A security monitoring system for Internet of Things. *Internet of Things* 7:100080. <https://doi.org/10.1016/j.iot.2019.100080>
11. Vávra J, Hromada M, Lukáš L, Dworzecki J (2021) Adaptive anomaly detection system based on machine learning algorithms in an industrial control environment. *Int J Crit Infrastruct Prot* 34, 100446. <https://doi.org/10.1016/j.ijcip.2021.100446>
12. Wang C, Wang B, Liu H, Qu H (2020) Anomaly detection for industrial control system based on autoencoder neural network. *Wirel Commun Mob Comput* 2020:1–10. <https://doi.org/10.1155/2020/8897926>
13. Parra G, Rad P, Choo K-KR, Beebe N (2020) Detecting Internet of Things attacks using distributed deep learning. *J Netw Comput Appl* 163:102662. <https://doi.org/10.1016/j.jnca.2020.102662>
14. Cepheli Ö, Büyükkorak S, Karabulut Kurt G (2016) Hybrid Intrusion Detection System for DDoS Attacks. *J Electr Comput Eng* 2016. <https://doi.org/10.1155/2016/1075648>
15. Li S, Tryfonas T, Li H (2016) The Internet of Things: a security point of view. *Internet Res* 26:337–359. <https://doi.org/10.1108/IntR-07-2014-0173>
16. Das S, Namasudra S (2022) A lightweight and anonymous mutual authentication scheme for medical big data in distributed smart healthcare systems. *IEEE/ACM Trans Comput Biol Bioinforma* 1–12. <https://doi.org/10.1109/TCBB.2022.3230053>
17. Das S, Namasudra S (2023) Lightweight and efficient privacy-preserving mutual authentication scheme to secure Internet of Things-based smart healthcare. *Trans Emerg Telecommun Technol* 1–15. <https://doi.org/10.1002/ett.4716>

18. Wu F, Li X, Xu L, Kumari S, Karuppiah M, Shen J (2017) A lightweight and privacy-preserving mutual authentication scheme for wearable devices assisted by cloud server. *Comput Electr Eng* 63:168–181. <https://doi.org/10.1016/j.compeleceng.2017.04.012>
19. Huang Y-K (2023) Design of a smart cabin lighting system based on internet of things. *Cloud Comput. Data Sci* 112–121
20. Namasudra S (2022) A secure cryptosystem using DNA cryptography and DNA steganography for the cloud-based IoT infrastructure. *Comput Electr Eng*. [https://doi.org/10.1016/j.compel\\_eceng.2022.108426](https://doi.org/10.1016/j.compel_eceng.2022.108426)

# A Hybrid Peer-to-Peer Data Center Resource Management System



Shreyas S. Kaundinya, S. Shreyas, Joel Macklyn Dsouza,  
K. Gagan Prashanth, and Prafullata K. Auradkar

**Abstract** Data centers are extensively used in computer science research and development, in varying degrees for performing computations that require more power or easier access than personal computers can ever provide. Cloud service providers organize their infrastructure into multiple such data centers. These resources are effectively managed and utilized through several levels of automation. The provision of multiple such on-demand services catered by these clusters makes it a particularly challenging task for system administrators to monitor and manage these data centers. Contemporary tools and technologies for overall resource management are either commercially available or provide isolated modules in the compute monitoring and management pipeline. The temporal variations of workloads in academic institutions/data centers are huge as compared to any of the large-scale data centers of real-world businesses. Thus, it requires an efficient monitoring and management system to match the varying resource demands with the finite capacity infrastructure. This paper addresses the challenges through (i) virtualized resource allocation and monitoring (ii) an end-end pipeline with open-source tools, tailored for small and medium-scale enterprises and institutions (iii) effective visualization of the data center resource utilization.

**Keywords** Virtual Machines (VM) · Server management · VM orchestration · Dynamic allocation · System monitoring

## 1 Introduction

Data centers are crucial in today's digital landscape, providing cost-effective solutions for storing and managing large amounts of data and offering powerful computational resources. These resources, often available as on-demand services, allow users

---

S. S. Kaundinya (✉) · S. Shreyas · J. Macklyn Dsouza · K. Gagan Prashanth · P. K. Auradkar  
Department of CSE, PES University, Bengaluru 560085, India  
e-mail: [shreyassk08@gmail.com](mailto:shreyassk08@gmail.com)

P. K. Auradkar  
e-mail: [Prafullatak@pes.edu](mailto:Prafullatak@pes.edu)

to scale up or down as needed. However, as user and job numbers increase, managing these resources becomes increasingly challenging for system administrators, especially in tracking available resources and ensuring efficient allocation.

Large institutions have in-house solutions for resource management which are proprietary tools, such as Google's Borg [1], a cluster manager capable of running hundreds of thousands of jobs across numerous clusters. Borg offers features like process-level performance isolation, machine sharing, and reduced fault recovery time. However, these tools are tailor-made for larger institutions, making them difficult to adapt and deploy in small to medium-sized organizations, such as research centers and universities, which typically operate with budgets not exceeding 5–10 crore INR [2].

To address these challenges, various techniques and tools have been developed. Virtualization technology partitions physical resources into smaller, manageable units that can be allocated on demand while ensuring security through isolation. Load-balancing algorithms distribute workloads across multiple servers, optimizing resource use and preventing server overload. Energy-efficient strategies like dynamic voltage and frequency scaling (DVFS) optimize power usage without compromising performance, reducing both environmental impact and operational costs.

Despite these advancements, implementing industry-standard solutions in small to medium-sized data centers remains challenging due to the limited hardware resources and expertise required for setup, deployment, and management.

This paper introduces a Data Center Resource Management system aimed at small to medium-sized centers, such as those in colleges and research centers. The proposed solution employs a hybrid peer-to-peer architecture to efficiently handle user requests for compute resource allocation via Virtual Machines, ensuring security through isolation. This distributed architecture enhances fault tolerance, supports numerous user requests, and facilitates effective monitoring and management of services within the data center.

The main contributions of the paper are as follows:

1. Development of a data center management system tailored for small to medium-sized data centers.
2. Adoption of a secure and efficient resource allocation strategy.
3. Simplify job of system administrators through an effective visualization of resources.

The rest of the paper is organized as follows: Sect. 2 provides an overview of related work in the field. Sections 3 and 4 discuss the implementation details of the project. The remaining sections present the results, conclusions, and potential future expansions of the project.

## 2 Related Work

With the establishment of more data centers, there is a need for a system that can efficiently manage compute clusters and scale easily. Extensive research has led to the development of various products with these capabilities. However, the primary reliance on centralized architectures, while facilitating management, often limits the extent of scalability due to a centralized bottleneck. OpenStack Compute [3] is an open-source tool that is widely used cloud computing software that provides comprehensive self-service environments for sharing and managing compute, network, and storage resources in the data center. This however uses a centralized web-based portal. This led to further studies that aimed to de-centralize these systems.

One such type of architecture is that of a Peer-to-Peer(P2P) system [4]. P2P systems provide a distributed approach to tackling this, thus providing more scalability, but they are limited by internal communications. However, research has been done in order to optimize the load on data centers while handling tasks following a P2P model [5].

Research in this domain is mainly focused on integrating Peer-to-Peer mechanisms into data center operations to assess their viability in constructing Cloud Computing infrastructures and weighing their advantages and limitations [6–9]. These studies tackle the fundamental issue of maintaining system coherence in the face of unreliable computing resources. Moreover, they showcase a Java prototype implementation, demonstrating the practicality of a completely decentralized P2P Cloud framework. This set of papers provides proof for the implementation of Peer-to-Peer architecture in data centers.

Further studies led to research that focused on different algorithms for resource allocation and load balancing. Efficient Load Optimization and Resource Minimization (ELORM) [5] algorithm is used for optimizing the job load of different regions of data centers and users in a cloud environment. The method mainly concentrates on optimizing the load with minimal resource utilization task computation time and virtual machine cost. Another such algorithm distributed a local search algorithm on a population of peers, where each peer is considered as a potential solution checking its neighbors in the hope of finding an improved solution. "Neighbour Assisted Distributed and Scalable Environment" (NADSE) is an approach used to enhance performance compared to systems like Gnutella and Freenet [10].

Work done by Wuhib et al. [11] uses a gossip protocol to manage resources in clouds dynamically. Their approach involves nodes exchanging small messages with a subset of other nodes. Through these messages, nodes share state information to compute a new setup, aiming to maximize cloud performance. If the benefits of a new setup outweigh the costs, nodes accept the change and update their local state. The proposed approach is different in terms of the purpose of resource allocation. The authors' main focus is on the fairness of allocation, while the aim here is to deliver requested resources without any restrictions or delay.

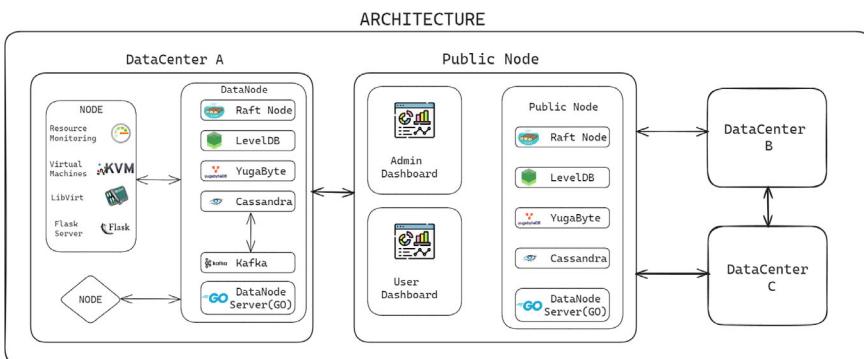
Throughout the survey into the intricacies of load-balancing algorithms, it was noticed that these algorithms demonstrate remarkable efficiency in their operations.

So the strategy is to pivot toward addressing other pertinent aspects of system design. This shift allows for a focus on building the remaining aspects of the system. The goal is to create a flexible framework capable of seamlessly accommodating a wide range of relevant algorithms. This approach adds scalability and enables experimentation with novel load-balancing techniques that may be more relevant to the task at hand.

Another challenge that has popped up due to the exponential, explosive growth of data centers in the 2010s is the very large amount of electrical power needed to run them and control the environment in which the servers run. The energy consumption of data centers alone will rise from 200 TWh in 2016 to 2967 TWh in 2030, extrapolating data keeping in mind the end of Moore's law [12]. The data centers require a lot of power to provide services, which increases CO<sub>2</sub> emissions and leads to a big carbon footprint. Many methods such as DVFS are VM migration algorithms to stack many VMs in one machine so that the others can be kept on standby. Recent advances in VM migration have introduced techniques involving master servers and slave servers, which outperform most older methods [13]. Katal et al. [14] have conducted a very detailed survey on this topic. However, the current work focuses on automating virtual machine construction and resource monitoring, among other things. Energy efficiency, live VM migration, and other optimization features can be taken up as future scope.

### 3 Proposed Work

The system consists of modules mainly, Data node/Public node server which are the main controllers of the system, Node services which manage and maintain the systems at the data center and networking interfaces which bind all the systems together and let them communicate (Fig. 1).



**Fig. 1** Architecture consisting of multiple data centers

### 3.1 Components

The architecture presented is that of a Hybrid peer-to-peer model of a data center resource management system. This diagram shows an overview of the components present. The various components are as follows:

- **Datanode Server:** The datanode server is a module that is hosted on one of the machines within the data center, which acts as a controller for the data center. The responsibilities of the datanode server are as follows:
  - **Raft Node:** Raft consensus algorithm [15] provided the necessary mechanism to broadcast and accept proposals for Virtual Machines all while maintaining consensus among multiple datanodes in a multi-data center setup. This component also involves setting up and running an instance of LevelDB, a fast embedded Key-Value store used by the Raft instance to maintain the state and metadata of the cluster and the VM allocations.
  - **Yugabyte:** Yugabyte, a distributed SQL database is instantiated in each datanode in a data center. Raft is used as the underlying consensus algorithm to replicate the state among the geologically distributed nodes of Yugabyte. This database hosts the tables required to back the dashboards. Dashboards can be locally hosted in data centers. The ACID-compliant database of choice will help in providing a consistent view across different data centers. This allows the system to be managed by admin members across different data centers.
  - **Cassandra:** Each datanode additionally runs an instance of Cassandra, a distributed NoSQL database. Each instance of Cassandra consumes and stores the resource metrics of the nodes in its cluster. These resource metrics are constantly pushed via a process in the nodes, periodically to a Kafka instance. The metrics are then consumed by a custom Python Kafka Cassandra connector within the data node and pushed to Cassandra. The updated metrics are used to plot a live graph of individual system utilization, which can be viewed by the admin on the frontend and acts as a visual aid for monitoring the data center for any irregularities.
  - **DHCP server:** A DHCP server to hand out and manage IP addresses to newly created VMs in the server.
- **Node:** Nodes are the individual systems in the data center that get assigned the tasks of provisioning, terminating, or modifying virtual machines by the datanode server. The various components and functionalities of the node are as follows:
  - **Flask Server:** Each node hosts and runs a Flask server which acts as the main point of communication between the datanode server and node with respect to Virtual Machine operations. The flask server has various endpoints and respective actions for operations such as creating, deleting, and modifying virtual machines using LibVirt APIs.
  - **Resource Monitoring Daemon process:** Each node runs a background shell script which executes an executable written in Go and starts up a Flask server

along with the systems process. Go executable polls the system metrics such as RAM utilization, Disk utilization and CPU utilization, and writes these metrics to a topic on the Kafka broker which is then consumed by a Python connector written to sync changes into Cassandra for persistence. Flask server is essential in maintaining the life cycle of a Virtual Machine at the node level. It uses the Python wrapper for Libvirt [16] APIs to perform CRUD operations on virtual machines, maintain the networking capabilities, and allocate the required space to run the VM.

- **Libvirt API:** To create, delete, or modify virtual machines, the solution uses Libvirt [16], which is an open-source API, daemon, and management tool for managing platform virtualization. For the proposed solution, KVM virtualization [17] is used due to it being based on a Type-1 hypervisor and using the Linux process scheduler for efficient scheduling of the various calls. To make the overall process more seamless, a Libvirt API for Python was used since the API calls can be integrated with the responses of the Flask server explained above.
- **Network Interfaces:** The installation script for each node creates a virtual Linux network bridge in the node machine, which is a data-link layer (Layer 2) device that is bound to the physical ethernet interface. Linux bridge provides satisfactory performance for the intended purposes [18]. Each virtual machine hosted by the node machine can now connect directly to the data center’s network. Due to virtualization-isolation requirements, a Virtual Machine cannot be assigned an IP address by the host OS, however, this can easily be fixed by configuring a DHCP server on the datanode machine of each datacenter to hand out IP leases to them. This type of configuration allows more control of the datanode server—IPs can now be bound to MAC Addresses, and so static IPs can be handed out more easily when required without any communication or logic on the end of the node machines.
- **Disk Services:** A Linux service that runs on startup, which monitors the images folder. All VM disk images are contained in this directory. Upon creation of a new VM, one of the blank images is consumed by the VM and renamed. The service detects this and initiates the copy of a new blank VM from a repository, either locally or stored online, based on configuration.
- **Public Node Server:** The public node server acts as the central point of contact for administrators via a web interface. Similar to the datanode server, this public node hosts YugaByte, Cassandra, Kafka Consumer, and a Raft instance. When a user submits a Virtual Machine operation request, the proposal is directed through the public node to the Raft leader node, VM request is then assigned to a specific data center using the Raft consensus algorithm based on availability. Additionally, the hosted web server offers an intuitive frontend interface with real-time system utilization graphs sourced from the corresponding datanode server.

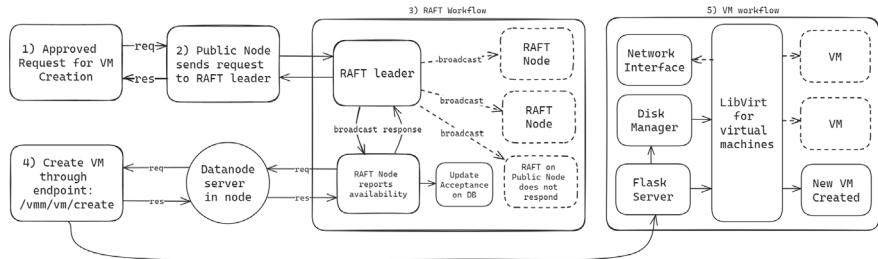


Fig. 2 Workflow for the process of allocating VM in a data center

### 3.2 Workflow

This hybrid approach leverages the strengths of traditional client-server architectures and the flexibility and robustness of P2P networks. The workflow is designed to optimize resource utilization, enhance fault tolerance, and ensure scalability.

Figure 2 details a workflow for one particular operation—create. The workflows for other operations are similar, though the Raft section is omitted in those cases.

The architectural framework comprises multiple components, each possessing individualized workflows and can be explained as below.

- **Datanode-Datanode communication:** The datanodes that represent the individual data centers in the network act as nodes in the Raft consensus process. Each datanode hosts a Raft node, a HTTP server, Kafka cluster, a Yugabyte Node, and a Cassandra Node. A Public Node receives the requests and relays them through the leader node (determined by the Raft leader election process) to all the datanodes, where the requests are replicated using Raft. Based on the availability of resources, a data center sends an accept message for the given request, which is again replicated among the nodes, after which the Virtual Machine is created. Other HTTP requests required by frontend are serviced by the HTTP server written in Go.
- **Virtual Machine Operations workflow:** In order for virtual machine operations to be executed, the node needs to receive the appropriate commands from the datanode server. In order to handle these communications, an HTTP server is set up using Flask to be constantly running in order to receive and handle the requests. The HTTP server has appropriate Libvirt API calls which are called when the endpoint corresponding to a certain VM operation receives a request. Since it is essential for the Flask-HTTP server to be constantly running, a shell process which is running in the background pings the process periodically, to restart or fix any issues.

The overall workflow for various VM operations can be summarized as below.

- **Create Operation:** The create operation is assigned to a node based on the availability of compute resources. This involves a look-up on the level db database on receiving the request and assignment of the request to a data center based on the

consensus derived from the Raft process. After that, an HTTP request is sent to the node that is assigned to handle the request with the user-given specifications of the VM to be created. The HTTP server in the node handles the request with the appropriate Libvirt calls mapped to the endpoints and on successful completion of the operation, it sends the updated system metrics to the datanode server via Kafka, which on receiving updates the values in the database. This operation is done as soon as the process of creation is completed in order to accurately handle subsequent requests. On failure, it sends a response back to the datanode server and actions can be taken accordingly.

- **Delete and update Operations:** The overall workflow for the delete and update operation is similar to that of the create operation. The major difference is that these operations involve direct communication without the execution of Raft. The node receives a delete or update request on its HTTP server with the Virtual machine and node parameters on which the operation is to be performed, and as a response, the appropriate Libvirt call is executed. On successful execution of the operation, the updated system metrics are sent to the datanode server.
- **Start, Suspend, and Shut Down operations:** The execution of the start, suspend, and shut down operations involves the datanode server sending a request to the HTTP server running in the node with the appropriate parameters of the virtual machine on which the operation is to be performed upon. From the user or admin perspective, this operation is done by clicking on the appropriate operation from the front end hosted on the public node server. These operations do not involve updating the system metrics since the overall system metrics are unchanged.
- **Backup Operation:** The implementation of the backup operation serves a dual purpose: to enhance system reliability and provide users with enhanced customization options when creating virtual machines. By incorporating a backup feature, the system aims to minimize the risk of data loss and streamline the process of provisioning virtual machines to meet specific user requirements. This functionality empowers end users to suspend existing services and efficiently load new virtual machines with the properties and configurations of previously created instances, thereby facilitating seamless transitions and reducing downtime. The various operations encompassed within the backup process can be summarized as follows:
  - **Create backup:** Spawns a thread that suspends the operation of a particular VM and creates a copy of its disk file in a separate backup directory. By default, up to three backups can be made, exceeding which the oldest existing backup will be rewritten. This number can be easily changed by changing a variable and re-installing the service script.
  - **Load backup:** Suspends the operation of a VM and replaces its current disk file with the selected backup. The current state of the VM can also be saved at the discretion of the user, however, this might delete the oldest existing backup.
  - **VM backup management:** Backed-up images can be downloaded by the user whenever necessary. VM images can be deleted as required.

## 4 Performance Analysis

The performance analysis section delves into the evaluation of the system's behavior under various conditions. There are two main sections, one describing the hardware setup used to test the system and the other discussing the results inferred from the testing.

### 4.1 Experimental Analysis

This section elaborates on the compute server setups utilized for conducting the experiments. It emphasizes that all server machines operated on Ubuntu 22.04 LTS, a choice made to maintain uniformity and stability throughout the testing environment. Additionally, each machine was interconnected with the internet and the datacenter network via dual 10GiB Ethernet Network Interface Cards (NICs), ensuring robust connectivity and high-speed data transfer capabilities.

#### 4.1.1 VM Workflows

VM workflows have been tested on machines with specifications:

- **CPU:** Intel Xeon(R) E3-1220 V2 server-grade CPU running at a base clock of 3.10GHz and a boost clock of 3.5GHz, with 4 threads.
- **RAM:** 16 Gigabytes DDR3 RAM.

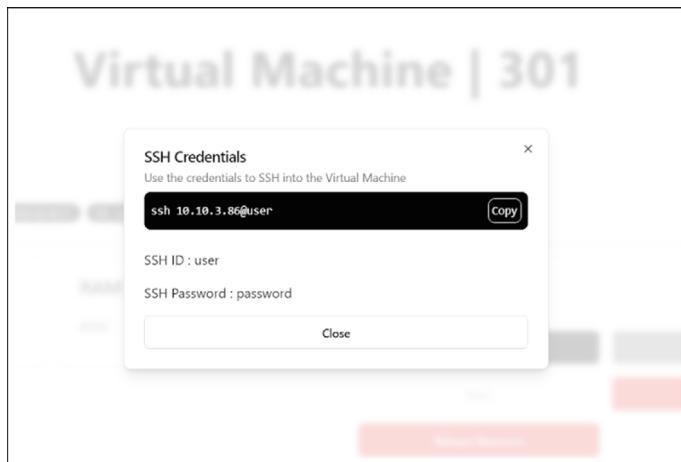
#### 4.1.2 Resource Usage Metrics

Resource metrics monitors were tested on an identical setup to host the VMs and a laptop machine that hosted the central datanode, running in a Docker (v24.0.6) container. This laptop had:

- **CPU:** AMD Ryzen(R) 5 4800H CPU running at a base clock of 3 GHz and a boost clock of 4 GHz, with 12 threads.
- **RAM:** 16 Gigabytes DDR3 RAM.

#### 4.1.3 Stress Testing

Key performance metrics, like throughput and latency, were measured to provide a comprehensive understanding of the system's behavior under stress. These tests were performed on the same workbenches as the other tests.



**Fig. 3** Front end screenshot of response on successful VM creation

## 4.2 Results and Discussion

This subsection deals with the findings from the three major experiments: VM workflow, Resource metrics, and stress tests performed on the system. Each of these experiments were aimed at different aspects of the system to provide insight into the performance of the system.

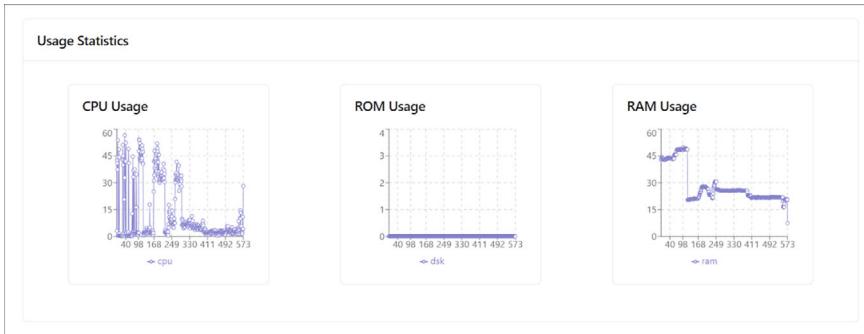
### 4.2.1 VM Workflows

During testing it was noted that it took about  $20 \pm 5$  s for a VM to be allocated from the moment the create request was approved by the admin and sent to the network. This behavior was consistently observed across the tests performed, leading to the implementation of a 30-second delay to ensure consistent behavior (Fig. 3).

The system successfully allocated hardware and network resources, and after doing so, it reported back with the IP of the created VM and login credentials for the VM.

### 4.2.2 Resource Usage Metrics

Resource usage metrics are observed on the frontend application, which is connected to the data node server. It was observed that the data recorded on both the VM host and the central server were identical when set at a polling rate of 10 s between data collections (Fig. 4).



**Fig. 4** Screenshot of the frontend dashboard showing metrics for a particular host. *ROM in the image refers to Disk Usage*

**Table 1** Stress testing results of various operations

Result	Threads	Connections	Req/sec		Latency (ms)	
			Avg	Max	Avg	Max
KV get	2	256	3890	5940	39.7	273.3
KV set	2	256	323.11	1140	447.78	916.59
KV get	1	32	6960	9140	4.99	42.06
KV set	1	32	606.96	990	53.79	213.87
Go	4	64	2890	5840	7.27	83

Note The various operations are as follows:

- KV Get: Retrieval of values from the Raft Node's Key-Value Store via a GET request
- KV Set: Inserting of values into the Raft Node's KV Store via a POST request
- Go: HTTP requests sent to the Go server

#### 4.2.3 Stress Testing

Key performance metrics, like throughput and latency, were measured to provide a comprehensive understanding of the system's behavior under stress.

The below-tabulated results highlight the comparative performance of the system across different HTTP methods, shedding light on any variations in response times and throughput under varying workloads (Table 1).

By varying thread counts and connection configurations, the exploration focused on a proof of concept for the effectiveness of the system handling concurrent requests and various operations within the environment.

## 5 Conclusions and Future Work

In conclusion, this paper addresses the challenges faced by small to medium-scale enterprises and institutions in managing their data centers efficiently. By proposing a hybrid peer-to-peer architecture, the system aims to seamlessly manage to compute resources and handle user requests for resource allocation via Virtual Machines, while facilitating effective monitoring and management of services within the data center. It presents a detailed breakdown of components and workflows, illustrating how each element contributes to the overall functioning of the system. Through stress testing and performance evaluations, this paper demonstrates a proof of concept for the effectiveness of the solution in handling various operations and tasks within the data center environment. With ongoing research and development, it could significantly enhance data center operations and advance cloud computing infrastructure.

Looking forward, several aspects of the current solution can be further investigated to optimize and expand its capabilities. One area of future work involves enhancing scalability and performance by exploring efficient resource allocation strategies or dynamic scaling policies to accommodate varying workloads and resource demands. Another promising direction is to improve energy efficiency and sustainability within data centers. This could involve implementing dynamic power management techniques to reduce operational costs and environmental impact while maintaining the quality of service. Using heuristic-based, energy-aware allocation optimizes resource provisioning to client applications, enhancing energy use while upholding the Quality of Service (QoS) [19]. Additionally, leveraging hardware heterogeneity to distribute workloads effectively can optimize energy consumption while sustaining operational quality [20]. Implementing VM migration policies will also be crucial for accommodating varying workloads and allocating resources more efficiently, providing benefits such as fault tolerance, load balancing, power management, and hardware management [21]. These future developments could significantly advance the capabilities and efficiency of small to medium-scale data centers.

## References

1. Verma A, Pedrosa L, Korupolu M, Oppenheimer D, Tune E, Wilkes J (2015) Large-scale cluster management at google with Borg. In: Proceedings of the tenth european conference on computer systems. EuroSys'15. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/2741948.2741964>
2. Working Group E (2007) Report of the working group on science & technology for small & medium scale enterprises (SMES)
3. Sefraoui O, Aissaoui M, Eleuldj M (2012) Openstack: toward an open-source solution for cloud computing. Int J Comput Appl 55(3):38–42
4. Fox G (2001) Peer-to-peer networks. Comput Sci Eng 3(3):75–77. <https://doi.org/10.1109/5992.919270>
5. Priya B, Gnanasekaran T (2020) To optimize load of hybrid p2p cloud data-center using efficient load optimization and resource minimization algorithm. Peer-to-Peer Netw Appl 13(2):717–728. <https://doi.org/10.1007/s12083-019-00795-3>

6. Babaoglu O, Marzolla M, Tamburini M (2012) Design and implementation of a p2p cloud system. In: Proceedings of the 27th annual ACM symposium on applied computing, pp 412–417
7. Kavalionak H, Montresor A (2012) P2p and cloud: a marriage of convenience for replica management. In: International Workshop on Self-Organizing Systems. Springer, pp 60–71
8. Ranjan R, Zhao L, Wu X, Liu A, Quiroz A, Parashar M (2010) Peer-to-peer cloud provisioning: Service discovery and load-balancing. *Cloud Comput: Principles, Syst Appl* 195–217
9. Tang C, Chang RN, So E (2006) A distributed service management infrastructure for enterprise data centers based on peer-to-peer technology. In: 2006 IEEE international conference on services computing (SCC'06). IEEE, pp 52–59
10. Patel R, Garg V (2011) Resource management in peer-to-peer networks: a Nadse approach. In: AIP conference proceedings, vol 1414. American Institute of Physics, pp 159–164
11. Wuhib F, Stadler R, Spreitzer M (2012) A gossip protocol for dynamic resource management in large cloud environments. *IEEE Trans Netw Serv Manag* 9(2):213–225
12. Koot M, Wijnhoven F (2021) Usage impact on data center electricity needs: a system dynamic forecasting model. *Appl Energy* 291:116798. <https://doi.org/10.1016/j.apenergy.2021.116798>
13. Gupta A, Namasudra S (2022) A novel technique for accelerating live migration in cloud computing. *Autom Softw Eng* 29(1):34
14. Katal A, Dahiya S, Choudhury T (2023) Energy efficiency in cloud computing data centers: a survey on software technologies. *Cluster Comput* 26:1845–1875
15. Ongaro D, Ousterhout J (2014) In search of an understandable consensus algorithm. In: 2014 USENIX annual technical conference (USENIX ATC 14). USENIX Association, Philadelphia, PA, pp 305–319. <https://www.usenix.org/conference/atc14/technical-sessions/presentation/ongaro>
16. Bolte M, Sievers M, Birkenheuer G, Niehörster O, Brinkmann A (2010) Non-intrusive virtualization management using libvirt. In: 2010 design, automation & test in europe conference & exhibition (DATE 2010), pp 574–579. <https://doi.org/10.1109/DAT.2010.5457142>
17. Kivity A, Liguori A (2007) KVM: the Linux virtual machine monitor. <https://api.semanticscholar.org/CorpusID:2408450>
18. Yu J (2004) Performance evaluation of Linux bridge
19. Beloglazov A, Abawajy JH, Buyya R (2012) Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. *Future Gener Comput Syst* 28:755–768
20. Zapater M, Ayala JL, Moya JM (2012) Leveraging heterogeneity for energy minimization in data centers. In: 2012 12th IEEE/ACM international symposium on cluster, cloud and grid computing (CCGRID 2012), pp 752–757. <https://doi.org/10.1109/CCGrid.2012.34>
21. Le T (2020) A survey of live virtual machine migration techniques. *Comput Sci Rev* 38:100304. <https://doi.org/10.1016/j.cosrev.2020.100304>

# Author Index

## A

- Adamu, Bashir Zak, 367  
Ali, Usman, 109  
Anitha, G., 287  
Anupam Sarma, 3  
Anurag De, 259, 329  
Apoorva Patel, 53  
Ashish Kumar, 53  
Ashwini, K., 81

## B

- Ballesteros, Luis, 313  
Battula Lalitesh, 329  
Bermúdez, Sergio, 313  
Bharathi, V., 219  
Bhuvanya, R., 203  
Binsu C. Kovoor, 27  
Birendra Biswal, 163

## C

- Chandralekha, M., 67  
Corredor, Oscar, 313

## D

- Das, Abhrajit, 147  
Debaditya Ghosh, 133  
Debopam Dey, 247  
Deepak Kumar Khandelwal, 233  
Dhruv Agrawal, 353  
Dwijen Rudrapal, 247

## E

- Ertam, Fatih, 367

## F

- Forhad, Shamim, 339

## G

- Gagan Prashanth, K., 383  
Gaurav Sharma, 95  
Gautam Pal, 259  
Gayathri P Salian, 15  
Gonzalez-Crespo, Ruben, 313  
Gulzar, Yonis, 109  
Gurmeet Kaur, 53

## H

- Hasan, Khandakar Kamrul, 339  
Hassan, Samia Binta, 339  
Havirbhavi, P., 81

## I

- Idris, Idris Ya'u, 109  
Imon Mukherjee, 41  
Ishita Mehta, 179  
Islam, S. M. Nahidul, 339

## J

- Jeet Nandigrami, 133  
Jhalak Dutta, 133  
Joel Macklyn Dsouza, 383  
Joshua, J., 329  
Josiah Samuel Raj, J., 287  
Jothy, N., 219

**K**

- Kalakanti Pawan Tej, 259  
 Karnam Shyam, 259  
 Kartik Gupta, 179  
 Kilincer, Ilhan Firat, 367  
 Kowju Gayatri, 163  
 Krupa Mehta, 123  
 Kujani, T., 203

**M**

- Mahesh Chandra Govil, 233  
 Manasa K Rao, 15  
 Manibharathi, D., 191  
 Mansam Wajira, 301  
 Matheswaran, P., 203  
 Mayank Kashyap, 53  
 Montenegro-Marin, Carlos Enrique, 313  
 Mugesh, C., 329

**N**

- Neelam Jain, 123  
 Ningthoujam Juleina, 301  
 Noor, Mohammad Ashiqur, 339  
 Nova, Sunjida Mushfiq, 339

**O**

- Ortiz, Jannet, 313

**P**

- Prafullata K. Auradkar, 383  
 Priyadarshini Jayadurga, N., 67  
 Priya Singh, 95

**R**

- Rabul Saikia, 3

Ranjan Jana, 41

Rashmi, M., 15  
 Rohit Kumar Dey, 133  
 Roopam Deka, 3

**S**

- Saikat Bandopadhyay, 133  
 Salam Shuleenda Devi, 3  
 Saleem, Kashif, 67  
 Sanjeev Kumar, 191  
 Sanjoy Debnath, 275  
 Santhosh Jayagopalan, 219  
 Saswati Debnath, 353  
 Shreyas, S., 383  
 Shreyas S. Kaundinya, 383  
 Siddique, Abdul Hasib, 339  
 Smita Das, 133  
 Sorokhaibam Nilakanta Meitei, 301  
 Soubam Chitra Devi, 301  
 Soumyajit Datta, 133  
 Sudha, T., 219  
 Sunu Fathima, T. H., 27

**T**

- Tanusha, G., 81  
 Tasnim, Jarin, 339

**V**

- Vasanthanayaki, C., 191  
 Vedavati Patil, 353  
 Vibhuthi Amarnath, 275

**Y**

- Yaddanapudi Srilekha, 275  
 Yaddanapudi Venkata Sri Harsha, 275  
 Ya'u, Badamasi Imam, 109