

Studies in Computational Intelligence 1145

Fausto Pedro García Márquez ·  
Akhtar Jamil · Isaac Segovia Ramirez ·  
Süleyman Eken ·  
Alaa Ali Hameed *Editors*

# Computing, Internet of Things and Data Analytics

Selected papers from the International  
Conference on Computing, IoT and Data  
Analytics (ICCIDA)

 Springer

Series Editor

Janusz Kacprzyk, *Polish Academy of Sciences, Warsaw, Poland*

The series “Studies in Computational Intelligence” (SCI) publishes new developments and advances in the various areas of computational intelligence—quickly and with a high quality. The intent is to cover the theory, applications, and design methods of computational intelligence, as embedded in the fields of engineering, computer science, physics and life sciences, as well as the methodologies behind them. The series contains monographs, lecture notes and edited volumes in computational intelligence spanning the areas of neural networks, connectionist systems, genetic algorithms, evolutionary computation, artificial intelligence, cellular automata, self-organizing systems, soft computing, fuzzy systems, and hybrid intelligent systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

Indexed by SCOPUS, DBLP, WTI Frankfurt eG, zbMATH, SCImago.


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

Fausto Pedro García Márquez · Akhtar Jamil ·  
Isaac Segovia Ramirez · Süleyman Eken ·  
Alaa Ali Hameed  
Editors

# Computing, Internet of Things and Data Analytics

Selected papers from the International  
Conference on Computing, IoT and Data  
Analytics (ICCIDA)

*Editors*

Fausto Pedro García Márquez   
Ingenium Research Group, Campus  
Universitario  
University of Castilla-La Mancha  
Ciudad Real, Spain

Isaac Segovia Ramirez  
Ingenium Research Group, Campus  
Universitario  
University of Castilla-La Mancha  
Ciudad Real, Spain

Alaa Ali Hameed  
Department of Computer Engineering  
Istinye University  
Istanbul, Türkiye

Akhtar Jamil  
National University of Computer  
and Emerging Sciences  
Islamabad, Pakistan

Süleyman Eken  
Information Systems Engineering  
Kocaeli University  
Kocaeli, Türkiye

ISSN 1860-949X

ISSN 1860-9503 (electronic)

Studies in Computational Intelligence

ISBN 978-3-031-53716-5

ISBN 978-3-031-53717-2 (eBook)

<https://doi.org/10.1007/978-3-031-53717-2>

© The Editor(s) (if applicable) and The Author(s), under exclusive license  
to Springer Nature Switzerland AG 2024

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 Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

# Preface

This book is a compilation of the selected papers presented at the International Conference on Computing, IoT and Data Analytics (ICCIDA) in 2023.

The conference was hosted by Ciudad Real, Spain and organized by Ingenium Research Group at the University of Castilla-La Mancha, Spain, on July 20–21, 2023. The conference served as an interdisciplinary forum that helped the research community to take a step forward and share the research findings. In addition, it provided an arena where researchers, scholars, professionals, students and academicians may be able to foster working relationships and gain access to the latest research results.

The book highlights some of the latest research advances and cutting-edge analysis of real-world case studies on computational intelligence, data analytics, IoT and applications from a wide range of international contexts. It also identified business applications and the latest findings and innovations in operations management and the decision sciences, for example:

## Data Analysis and Visualization

- Exploratory Data Analysis
- Statistical and Mathematical Modeling
- Business Intelligence
- Big Data Analysis
- Data Mining
- Cloud Computing Architecture and Systems
- ETL and Big Data Warehousing
- Business Intelligence
- Data Visualization
- Statistical Analysis

## Computer Vision

- Document Analysis
- Biometrics and Pattern Recognition
- Remote Sensing & GIS
- Medical Image Processing
- Image and Video Retrieval
- Motion Analysis
- Structure from Motion
- Object Detection and Recognition
- Image Restoration
- Speech and Audio Processing
- Signal Processing

## Artificial Intelligence

- Machine Learning
- Pattern Recognition
- Deep Learning
- Human–Computer Interactions
- Medical Image Processing
- Image and Video Retrieval
- Audio Video Processing
- Text Analytics
- Natural Language Processing
- Information Retrieval
- Robotics Applications

## Internet of Things

- 3D Printing
- Securing IoT infrastructure
- Future of IoT and Big Data
- Internet of Things
- Intelligent Systems for IoT
- Security, Privacy and Trust
- Visual Analytics IoT
- Data Compression for IoT Devices
- IoT Services and Applications
- Education and Learning
- Social Networks Analysis

## Communication Systems and Networks

- Antennas, Propagation and RF Design
- Transmission and Communication Theory
- Wireless/Radio Access Technologies
- Optical Networks and NGN
- 5G & 6G Cellular Systems and SON
- Sensor Networks
- Multimedia and New Media
- High-Speed Communication.
- Computational Intelligence in Telecommunications

## Software Engineering

- Requirements Engineering
- Security Aspects
- Agile Software Engineering
- Software Evolution & Reuse
- Reverse Engineering
- Software Dependability
- Data & AI Monetization and Products

- Data-as-a-Service/Platform
- Biomedical Experiments and Simulations
- Decision Support Systems

Fausto Pedro García Márquez  
Akhtar Jamil  
Isaac Segovia Ramirez  
Süleyman Eken  
Alaa Ali Hameed



# Organization

## General Chairs

Fausto Pedro Garcia Marquez      University of Castilla-La Mancha, Spain  
Alaa Ali Hameed      Istinye University, Turkiye

## Program Chairs

Mehmet Alper Tunga      Istinye University, Turkiye  
Hatice Gülen      Istinye University, Turkiye

## Technical Program Chairs

Süleyman Eken      Kocaeli University, Turkiye  
Akhtar Jamil      National University of Computer and Emerging  
Sciences, Pakistan

## Publication and Publicity Chairs

Serdar Solak      Information Systems Engineering, Kocaeli  
University, Turkiye  
Mustafa Hikmet Bilgehan Uçar      Information Systems Engineering, Kocaeli  
University, Turkiye

## Organizing Committee

Amir Seyyedabbasi      Istinye University, Turkiye  
Kiran Sood      Chitkara University, India  
Faezeh Soleimani      Ball State University, USA  
Momina Mustehsan      Bahria University, Pakistan  
David Yeregui Marcos      University of Leon, Spain

## Registration Committee

Muhammad Davud	Istinye University, Turkiye
Enkeleda Lulaj	University Haxhi Zeka, Kosovo
Tianle Zhang	University of Exeter, UK

## Technical Program Committee

Ahmet Ali Suzen	Isparta University of Applied Sciences, Turkiye
Ahmet Cevahir	Selçuk University, Turkiye
Aytuğ Onan İzmir Katip	Çelebi University, Turkiye
Azidine Guezzaz	Technology Higher School Essaouira, Morocco
Bharat Bhushan	Sharda University, India
Caner Filiz	Baskent University, Turkiye
Doğan Aydın	Izmir Katip Celebi University, Turkiye
Durmus Koç	Usak University, Turkiye
G. Jayanthi	Sri Ramachandra Institute of Higher Education and Research, India
Gabriella Casalino	Università degli Studi di Bari Aldo Moro Bari, Italy
Gokhan Bakal	Abdullah Gul University
Halit ÖZTEKİN	Sakarya Uygulamalı Bilimler, Turkiye
Hsi-Ming Ho	University of Sussex, UK
Hüseyin Aygün	Maltepe University, Turkiye
Isidro Peña García-Pardo	University of Castilla-La Mancha, Spain
Jamal Mabrouki	University Mohammed V in Rabat, Morocco
Jayanthi Ganapathy	Sri Ramachandra Institute of Higher Education and Research, India
Jonathan Browning	Queen's University Belfast, UK
Lu Bai	Ulster University, UK
Mehak Khan	Oslo Metropolitan University, Oslo, Norway
Mehmet Bilen	Suleyman Demirel University, Turkiye
Melike Sah Direkoglu	Near East University, UK
Momina Mustehsan	Bahria University, Pakistan
Mourade Azrou	University Moulay Ismail of Meknes, Morocco
Muhammad Abdul Basit	Montana Technological University, Butte Montana, USA
Muhammed Karaaltun	Necmettin Erbakan University, Turkiye
Muhammed Karaaltun	Erbakan University, Turkiye
Mustafa Al-asadi	KTO Karatay University, Turkiye
Mustafa Takaoğlu	TÜBİTAK-BİLGEM, Kocaeli, Turkiye

Osama M. F. Dawoud	Istinye University, Turkiye
Özerk Yavuz	Bilkent University, Turkiye
Pijush Kanti Dutta Pramanik	Galgotias University Greater Noida, India
Said Benkirane	Technology Higher School Essaouira, Morocco
Selçuk Kavut	Balikesir University, Turkiye
Semih ÇAKIR	Zonguldak Bülent Ecevit University, Turkiye
Shaaban A.I. Sahmoud	Fatih Sultan Mehmet Vakif University, Turkiye
Shakir Khan	Imam Mohammad Ibn Saud Islamic University, Saudi Arabia
Subhan Ullah	National University of Computer and Emerging Sciences, Pakistan
Tuncay Ercan	Yasar University, Turkiye
Turgay İbrikci	Adana Alparslan Turkes Science and Technology University, Turkiye
Turgay İbrikci	Cukurova University, Turkiye
Turgay Tugay Bilgin	Bursa Teknik Üniversiteleri, Turkiye
Vijay Anant Athavale	Walchand Institute of Technology, India
Wadhah Zeyad Tareq	Istinye University, Turkiye
Waleed Ead	Beni-Suef University, Egypt
Yao Lu	Lancaster University, UK
Zhang, Tianle	University of Exeter, UK

# Contents

Artificial Intelligence Techniques in Precision Marketing: A Multi-criteria Analysis and Comparative Study .....	1
<i>Nouhaila El Koufi and Abdessamad Belangour</i>	
Artificial Intelligence in Industrial Internet of Things: A Concise Review of Performance Management .....	8
<i>Seda Balta Kaç and Süleyman Eken</i>	
Comparison of LDA, NMF and BERTopic Topic Modeling Techniques on Amazon Product Review Dataset: A Case Study .....	23
<i>Salih Can Turan, Kazım Yıldız, and Büşra Büyüktanir</i>	
Optimal Parameter Selection of Latent Dirichlet Allocation to Determine the Emerging Topics in Hydrology Domain .....	32
<i>Sila Ovgu Korkut, Aytug Onan, Erman Ulker, and Femin Yalcin</i>	
Battery Management System-Based Fuzzy Logic .....	43
<i>K. S. Jithin Mohan and S. Paul Sathiyar</i>	
Involvement of Unmanned Aerial Vehicles and Swarm Intelligence in Future Modern Warfare: An Overview .....	58
<i>Murat Bakirci and Muhammed Mirac Özer</i>	
A Comparative Study of Deep Learning Loss Functions: A Polyp Segmentation Case Study .....	68
<i>Rachid Bourday, Issam Aattouchi, and Mounir Ait Kerroum</i>	
Combination of an Improved Feistel Scheme and Genetic Operators for Chaotic Image Encryption .....	79
<i>Hicham Rrghout, Mourad Kattass, Mariem Jarjar, Naima Benazzi, Younes Qobbi, Abdellatif Jarjar, and Abdelhamid Benazzi</i>	
Post-processing of Closed Contours to Obtain Inscribed K-Sided Polygons .....	92
<i>R. Molano, M. Ávila, J. C. Sancho, P. G. Rodríguez, and A. Caro</i>	
Electricity Consumption Forecasting Using the Prophet Model in Industry: A Case Study .....	102
<i>Umut Yıldız and Sila Ovgu Korkut</i>	

Going Beyond Traditional Methods: Using LSTM Networks to Predict Rainfall in Kerala .....	112
<i>J. Akshaya, D. Harsha, D. Eswar Chowdary, B. E. Pranav Kumar, G. Rahul, V. Sowmya, E. A. Gopalakrishnan, and M. Dhanya</i>	
Effects of Augmented Reality on Visuospatial Abilities of Males and Females .....	122
<i>Julia Bend and Anssi Öörni</i>	
Parkinson's Disease Assessment from Speech Data Using Recurrence Plot .....	132
<i>Arnya Mohamed Ali, G. Jyothish Lal, V. Sowmya, and E. A. Gopalakrishnan</i>	
COVID-19 and Behavioral Analytics: Deep Learning-Based Work-From-Home Sensing from Reddit Comments .....	143
<i>M. M. Enes Yurtsever, Ekin Ekinci, and Süleyman Eken</i>	
An Improved Data Classification in Edge Cloud-Assisted IoMT: Leveraging Machine Learning and Feature Selection .....	156
<i>Abdelkarim Ait Temghart, Mbarek Marwan, and Mohamed Baslam</i>	
Predicting Delays in Indian Lower Courts Using AutoML and Decision Forests .....	166
<i>Mohit Bhatnagar and Shivaraj Huchhanavar</i>	
Chaotic Image Encryption Using an Improved Vigenère Cipher and a Crossover Operator .....	181
<i>Mourad Kattass, Hicham Rrghout, Mariem Jarjar, Abdellatif Jarjar, Faiq Gmira, and Abdelhamid Benazzi</i>	
NBS: An NFT-Based Blockchain Steganography Method .....	192
<i>Mustafa Takaoğlu, Faruk Takaoğlu, and Taner Dursun</i>	
MLP Neural Network Based on PCA and K-means Clustering for PM2.5 Forecasting .....	202
<i>Diego Velez, Santiago Santa, and Gustavo Patino</i>	
Comparing How Python and R Estimate Granger-Causality in the Frequency Domain .....	213
<i>Matteo Farnè and Meng Yang</i>	
Performance Analyses of AES and Blowfish Algorithms by Encrypting Files, Videos, and Images .....	223
<i>Karrar Hamzah Mezher and Timur Inan</i>	

A Wireless Emergency Alerts System for Warning Disasters by Using Distributed Databases, GPS and Machine Learning Enabled API Services . . . . .	231
<i>Md. Abdullah Al Mamun, Md. Tanvir Miah Shagar, Meher Durdana Khan Raisa, Md. Jubayer Hossain, Utsa Chandra Sutradhar, S. Rayhan Kabir, Anupam Hayath Chowdhury, and Mohammad Kamrul Hasan</i>	
Agent-Based Model for Oil Storage Monitor and Control System Using IoT . . .	241
<i>Hassan Kanj, Abdullah Aljeri, and Tarek Khalifa</i>	
A Novel Method to Detect High Impedance Fault in Electric Vehicle Integrated Distribution System . . . . .	251
<i>Pampa Sinha, S. Ramana Kumar Joga, Kaushik Paul, and Fausto Pedro García Márquez</i>	
Multi-class Classification of Voice Disorders Using Deep Transfer Learning . . .	262
<i>Mehtab Ur Rahman and Cem Direkoglu</i>	
Coronary Artery Blockage Detection by Automated Segmentation of Vessels in X-Ray Angiograms . . . . .	271
<i>Jayanthi Ganapathy, Fausto Pedro García Márquez, and C. H. Dhamini</i>	
HIYAM, A New Moroccan Humanoid Robot for Healthcare Applications Using IoT and Big Data Analytics . . . . .	281
<i>Hiba Asri and Zahi Jarir</i>	
Optimal Forecast Combination with Univariate Models for Natural Gas Prices in Spain . . . . .	288
<i>Roberto Morales-Arsenal, María Pilar Zazpe-Quintana, and Jesús María Pinar-Pérez</i>	
CNN for Efficient Objects Classification with Embedded Vector Fields . . . . .	297
<i>Oluwaseyi Igbasanmi, Nikolay M. Sirakov, and Adam Bowden</i>	
CEIMVEN: An Approach of Cutting Edge Implementation of Modified Versions of EfficientNet (V1-V2) Architecture for Breast Cancer Detection and Classification from Ultrasound Images . . . . .	310
<i>Sheekar Banerjee and Md. Kamrul Hasan Monir</i>	
Performance Improvement of Movie Recommender System Using Spectral Bi-clustering with Mahalabonis Distance . . . . .	324
<i>Sonu Airen and Jitendra Agrawal</i>	

Exploring the Use of Bluetooth Low Energy (BLE) Beacons for Enhancing Ecotrails in the Amazon Jungle of Peru ..... 334  
*J. Baldeón, D. Auccapuri, A. Masuda, R. Gálvez, E. Díaz, A. Arana, P. Chávez, V. Hernández, and M. Lau*

Identifying Success Factors in Business Intelligence Using Resource-Based Approach: A Literature Review ..... 344  
*Ruksana Banu*

Predictive Modeling of Breast Cancer Subtypes Using Machine Learning Algorithms ..... 354  
*Ashima Aggarwal and Anurag Sharma*

A Systematic Study on Unimodal and Multimodal Human Computer Interface for Emotion Recognition ..... 363  
*Akram Ahmad, Vaishali Singh, and Kamal Upreti*

Explainable AI Assisted Decision-Making and Human Behaviour ..... 376  
*Muhammad Suffian*

Performance Measurement of Classification Algorithms for Aerial Image Registration ..... 386  
*Hayder Mosa Merza, Ihab Sbeity, Mohamed Dbouk, Zein Al Abidin Ibrahim, and Ali Salam Kadhim*

Investigating IoT-Enabled 6G Communications: Opportunities and Challenges ..... 404  
*Radia Belkeziz, Reda Chefira, and Oumaima Tibssirte*

BlindEye: Blind Assistance Using Deep Learning ..... 420  
*Bilal Shabbir, Ali Salman, Sohaib Akhtar, and M. Asif Naeem*

Blockchain-Enhanced Privacy and Security in Medicare Data Sharing: Identifying Gaps and Solutions in Current Practices ..... 432  
*Abdullah Rehman and Muhammad Ilyas*

Enhanced Image Generation with MorphoGAN: Combining MNNs and GANs ..... 443  
*Islam M. Momtaz A. Sadek and Abdullahi Abdu Ibrahim*

Implementation of Homomorphic Encryption Schemes in Fog Computing ..... 455  
*Shrayash Pandey, Bharat Bhushan, Alaa Ali Hameed, Akhtar Jamil, and Aayush Juyal*

Automated and Optimised Machine Learning Algorithms for Healthcare Informatics ..... 465  
*Aayush Juyal, Bharat Bhushan, Alaa Ali Hameed, Akhtar Jamil, and Shrayash Pandey*

A Survey on Image-Based Cardiac Diagnosis Prediction Using Machine Learning and Deep Learning Techniques ..... 478  
*Anindya Nag, Biva Das, Riya Sil, Alaa Ali Hameed, Bharat Bhushan, and Akhtar Jamil*

Short Message Service Spam Detection System for Securing Mobile Text Communication Based on Machine Learning ..... 492  
*Ayasha Malik, Veena Parihar, Bharat Bhushan, Alaa Ali Hameed, Akhtar Jamil, and Pronaya Bhattacharya*

Object Identification: Comprehensive Approach Using Machine Learning Algorithms and Python Tools ..... 508  
*Mustafa Al-Asadi and Bharat Bhushan*

**Author Index** ..... 521



## About the Editors



**Fausto Pedro Garcia Marquez** Fausto works at UCLM as Full Professor (Accredited as Full Professor from 2013), Spain, **Honorary Senior Research Fellow at Birmingham University, UK, Lecturer at the Postgraduate European Institute**, Research Fellow at **INTI International University&Colleges**, Malaysia, and he has been Senior Manager in Accenture (2013–2014). He obtained his European PhD with maximum distinction. He has been **distinguished with the prizes**: Runner Prize (2023), Nominate Prize (2022), Gran Maestre (2022), Grand Prize (2021), Runner Prize (2020) and Advancement Prize (2018), Runner (2015), Advancement (2013), and Silver (2012) by the International Society of Management Science and Engineering Management (ICMSEM), First International Business Ideas Competition 2017 Award (2017), etc. **He has published more than 248 papers** (156 JCR: 74-Q1; 42-Q2; 32-Q3; 8-Q4), some recognized as: “Progress in Photovoltaics: Research and Applications” (Q1, IF. 8.49, one of the most downloaded in first 12 months of publications, 2023), “Applied Energy” (Q1, IF 9.746, as “Best Paper 2020”), “Renewable Energy” (Q1, IF 8.001, as “Best Paper 2014”); “ICMSEM” (as “excellent”), “Int. J. of Automation and Computing” and “IMechE Part F: J. of Rail and Rapid Transit” (most downloaded), etc. He is the author and editor of over **50 books** (Elsevier, Springer, Pearson, McGraw Hill, Intech, IGI, Marcombo, AlfaOmega,...), >100 international chapters, and six patents. He is the Editor of five Int. Journals, Committee Member more than 70 Int. Conferences. He has been the **Principal Investigator in four European Projects, eight National Projects, and more than 150 projects for universities, companies, etc.** His main interests are artificial intelligence, maintenance, management, renewable energy, transport, advanced analytics, and data science.

He is being: Expert in the European Union in AI4People (EISMD), and ESF.; Director of <https://www.ingenium.uclm.es/>; Senior Member at IEEE, 2021-... ; Honoured Honourary Member of the Research Council of Indian Institute of Finance, 2021-... ; Committee Chair of The International

Society for Management Science and Engineering Management (ISMSEM), 2020-.... His main interests are artificial intelligence, maintenance, management, renewable energy, transport, advanced analytics, data science.



**Dr. Akhtar Jamil** is an Associate Professor in the Department of Computer Science at the National University of Computer and Emerging Sciences, Islamabad, Pakistan. Before joining FAST, he served as Assistant Professor and Vice Head of the Computer Engineering Department at Istanbul Sabahattin Zaim University, Istanbul, Turkiye. He also served as a Lecture at COMSATS University Islamabad. He has also worked in the industry as a developer for several years. He received his Ph.D. in machine learning and remote sensing from Yildiz Technical University, Istanbul, Turkiye. He has published more than 50 high-quality papers in well-known journals and top conferences. He received a fully funded Ph.D. scholarship from the Turkish Government. He is the founding member of the [ICMI](#), [ICAETA](#), and [ICCIDA](#) conferences. He serves as a reviewer for several journals and conferences. He focuses on applied research for solving real-world problems. His current research interests include statistical machine learning, deep learning, pattern recognition, data analytics, image classification, and remote sensing.



**Isaac Segovia Ramirez** Industrial Engineer at ETSII of the University of Castilla-La Mancha, Ciudad Real (2015) and Master of Industrial Engineering at ETSII of the University of Castilla-La Mancha, Ciudad Real (2019). Associate professor of electronic courses in 2019–2020 in ETSII of the University of Castilla-La Mancha, Ciudad Real. Current PhD student, he is collaboration on several national and European projects with the Department of Business Administration of the University of Castilla-La Mancha and the (June 2013–present) Ingenium group. Winner of the international contest “Entrepreneurship 5+5” in Tunisia. He has been awarded with the “Advancement Prize for Management Science and Engineering Management with the Nominated Prize (2018)” for the article “Remotely Piloted Aircraft System and Engineering Management: A Real Case Study.” His main research interests are related to maintenance management, UAVs, renewable energy, detection of elements on surface by infrared radiation, etc.





**Süleyman Eken** received his MS degree and PhD degree in Computer Engineering at the Kocaeli University. He works as an Associate Professor of Information Systems Engineering, Kocaeli University, Izmit, Turkiye. He serves as a reviewer for more than 50 journals such as the IEEE Transactions on Industrial Informatics, JIS, Soft Computing, IEEE Access, International Journal of Advanced Robotic Systems, Concurrency and Computation: Practice and Experience, Journal of Ambient Intelligence and Humanized Computing, Imaging Science Journal, Journal of Network and Computer Applications, Turkish Journal of Electrical Engineering & Computer Sciences, Peer-to-Peer Networking and Applications. He has ten book chapters and around 150 papers. He served as guest editor for three issues. Also, he serves as associate editor for Cluster Computing (Springer) and Concurrency and Computation: Practice and Experience (Wiley).



**Alaa Ali Hameed** received his master's degree in computer engineering from Eastern Mediterranean University, North Cyprus, in 2012, and his Ph.D. degree from the Department of Computer Engineering at Selcuk University, Turkiye, in 2017. He worked as an Assistant Professor in the Department of Computer Engineering, at Istanbul Aydin University, Turkiye, from 2017 to 2019. He then moved to Istanbul Sabahattin Zaim University, Turkiye, where he worked as an Assistant Professor in the Department of Computer Engineering from 2019 to 2022. Currently, he is an Assistant Professor in the Department of Computer Engineering at Istinye University, Turkiye. He has published more than 60 technical articles in top international journals and conferences in a short span of time. He has served as a Program Chair and a Technical Program Chair member for many international conferences, and also he has served as a Guest Editor for many SCIE journals. His research interests include digital signal and image processing, adaptive filters, adaptive computing, data mining, machine, deep learning, big data and data analytics, neural networks and self-learning systems, and artificial intelligence.



# Artificial Intelligence Techniques in Precision Marketing: A Multi-criteria Analysis and Comparative Study

Nouhaila El Koufi<sup>(✉)</sup>  and Abdessamad Belangour 

Laboratory of Information Technology and Modeling, Faculty of Sciences Ben M'sik, Hassan II University, Casablanca, Morocco  
elkoufi.nouhaila1@gmail.com

**Abstract.** The incorporation of AI techniques has assumed a critical role in enhancing marketing strategies and ensuring sustained business growth. AI, armed with well-curated data and effective training, exhibits the capability to anticipate consumer preferences and deliver recommendations, thereby enabling precise marketing. Through the application of AI methods, precision marketing enables personalized interactions between businesses and customers. It attracts potential clients, offers tailored marketing recommendations to high-value customers, and ultimately reduces marketing costs. Thus, this paper aims to conduct a comprehensive analysis of the chosen methods using a predefined set of criteria. The comparative study will commence by introducing each method, followed by the selection of criteria for comparison. To obtain results, a multi-criteria comparative methodology has been adopted, which aligns perfectly with our objectives. The outcomes obtained from this evaluation will expose the strengths and weaknesses of each algorithm, thereby identifying areas for potential future improvements.

**Keywords:** Multi-criteria analysis · Precision Marketing · Comparative Study · Machine learning · AI

## 1 Introduction

In the marketing domain, the continuous evolution of artificial intelligence (AI) and digital technology stands as a transformative influence, offering numerous advantages to businesses and their marketing strategies. This synergistic connection between AI and digital technology has fundamentally reshaped the manner in which companies engage with their target audience, comprehend consumer behavior, and optimize their marketing endeavors. AI, known for its capacity to analyze vast datasets and discern patterns, equips marketers to make data-driven decisions with unparalleled precision. Through the utilization of machine learning algorithms, AI can sift through extensive data, extract valuable insights, and identify consumer trends and preferences. This empowers marketers to develop highly targeted and personalized campaigns, ensuring that the right message reaches the right audience at precisely the right moment.

Digital technology, on the other hand, provides the infrastructure and tools necessary to implement AI-driven marketing strategies effectively. With the proliferation of social media channels and Internet technology, businesses have unprecedented access to consumer data and touchpoints. Digital technologies allow marketers to collect and analyze data from various sources, gaining a comprehensive understanding of consumer preferences, behaviors, and interactions across multiple channels. The integration of AI and digital technology also enables marketers to optimize their campaigns and measure their effectiveness in real-time. Through advanced analytics and predictive modeling, businesses can continuously refine their marketing strategies, identifying areas of improvement and adjusting their tactics accordingly. This iterative approach allows for agility and adaptability, ensuring that marketing efforts are always aligned with changing consumer needs and market dynamics.

The remainder of this study is structured as follows: Sect. 2 presents and discusses the background, identifies comparison criteria, and in Sect. 3, it presents and discusses the advantages and drawbacks of the chosen techniques, along with the comparison results. Finally, a conclusion summarizing the exploration in this paper will be provided.

## 2 Background

### 2.1 Random Forest

Random Forest method stands as one of the most extensively employed techniques in the field of machine learning. Its inception can be attributed to Leo Breiman in the year 2001. This algorithm falls within the purview of supervised learning. The essence of the random forest method lies in its approach of constructing an ensemble of numerous decision trees, each of which is generated using two distinct sources of randomization. The initial form of randomization pertains to training each decision tree with a random sample. This sample is generated by selecting data points from the original dataset with replacement, maintaining the same size as the training set. The second source of randomness is associated with attribute sampling, where, in order to identify the best split at each node, a randomly chosen subset of input variables is utilized. The literature related to machine learning application in precision marketing domain, various studies employed random forest, like in [2] the authors have compared the performance of random forest and two others ML methods using a China banking that offer personal loan. The authors aimed to classify the customers in two classes. Yue Zhang [3] adopted logistic regression and random forest machine learning models and compared in terms of predicting consumer orders. The models achieve good accuracy in predicting orders, with random forest performing better in accuracy. The paper proposed by Zhu et al. [4] focuses on the rapid growth of behavioral data in tourism e-commerce and its importance for precision marketing. It conducts an empirical analysis of online purchases of tourism products to develop a set of effective features for predicting online purchases. The random forest method was adopted to build a predictive model.

## 2.2 SVM

Support Vector Machine (SVM) stands out as a widely adopted supervised learning technique introduced by Vapnik and colleagues in 1992. SVM is primarily employed for addressing binary and multiclassification issues [5]. SVM's central objective revolves around establishing a decision boundary, referred to as a hyperplane, which efficiently divides an n-dimensional space into distinct classes, allowing for the rapid categorization of new data points [6]. This hyperplane is strategically positioned to maximize its separation from the nearest data points in each class, known as support vectors. In the literature, various studies have harnessed the SVM method. For instance, in [7], authors utilized SVM to devise marketing strategies capable of effectively managing emergencies, promptly addressing risks, resolving business crises, and fostering rapid enterprise development. Similarly, [8] introduced a predictive model based on the SVM approach.

## 2.3 Artificial Neural Network

The artificial neural networks (ANN) method was introduced in 1944. ANN represent a collection of algorithms aimed at uncovering inherent connections within a dataset by emulating the functioning of the human brain. Their adaptability to changing inputs allows ANN to deliver optimal outcomes without necessitating alterations to the output criteria. Within an ANN, a neuron operates as a mathematical function responsible for gathering and classifying data following a specific structure. ANN function through three key layers: the input layer, which receives data, the hidden layer, which processes it, and the output layer, transmitting the computed results. Various studies were used ANN modeling in the literature [9, 10].

## 2.4 Comparison Criteria

- **Accuracy:** This criterion indicates the accuracy value of each model. It measures the proportion of accurately categorized instances from the overall number of instances. A model demonstrating high accuracy produces predictions that closely match the ground truth or anticipated results.
- **Speed of learning:** The speed of learning refers to how quickly a machine learning model can acquire knowledge or update its internal parameters based on the training data. It measures the efficiency of the learning process and determines how fast the model can reach a satisfactory level of performance. A model with a fast speed of learning can quickly adapt to new information and adjust its predictions accordingly.
- **Speed of classification:** This criterion refers to the efficiency of a machine learning model in making predictions or classifying new instances. It measures how quickly the model can process input data and provide an output.
- **Tolerance to missing values:** This criterion evaluates how well a machine learning model can handle instances or features with missing values. A model with high tolerance to missing values can still make accurate predictions even when some data is missing.
- **Tolerance to irrelevant attributes:** It assesses a model's capacity to disregard or reduce the influence of irrelevant features or attributes in the dataset. A model with

high tolerance to irrelevant attributes focuses on the most relevant features for making predictions.

- **Tolerance to redundant attributes:** This criterion assesses the model's capability to handle redundant features or attributes in the dataset. A model with high tolerance to redundant attributes can effectively identify and remove redundant information without negatively impacting its performance.
- **Tolerance to highly interdependent attributes:** It refers to the ability of a machine learning model to handle attributes that are highly dependent or correlated with each other. A model with high tolerance to highly interdependent attributes can still make accurate predictions even when faced with strong correlations between features.
- **Dealing with discrete/binary/continuous attributes:** This criterion examines how well a model can handle different types of attributes. Models should be capable of handling discrete (categorical), binary (two-value), and continuous (numeric) attributes appropriately, as these types of attributes require different processing techniques.
- **Tolerance to noise:** It measures how well a machine learning model can handle noisy or erroneous data. A model with high tolerance to noise can still produce reliable predictions even when the input data contains random or irrelevant fluctuations.
- **Dealing with danger of overfitting:** This criterion assesses a model's capability to prevent overfitting, a situation in which a model becomes excessively intricate and fits the training data too closely. This can result in poor generalization when applied to unseen data. A model that effectively prevents overfitting performs well when dealing with new, previously unseen instances.
- **Attempts for incremental learning:** This pertains to the model's capacity to acquire knowledge from new data in an incremental fashion, eliminating the need for a full retraining process. Incremental learning empowers the model to adjust and refresh its understanding continuously as new instances are introduced.
- **Explanation ability/transparency of knowledge/classification:** This criterion assesses the model's ability to provide explanations or interpretations for its predictions. Models with high explanation ability or transparency can provide insights into how and why certain predictions were made, making them more interpretable to humans.
- **Support Multiclassification:** This criterion examines whether a model can handle classification tasks with multiple classes or categories. Models that support multiclassification can effectively assign instances to more than two distinct classes or categories.

### 3 Discussion

Within the literature, numerous research endeavors have addressed and enhanced precision marketing by employing a range of artificial intelligence techniques. This study provides a comprehensive review of AI methodologies employed in PM. The selection of the most commonly used approaches within each AI method category was based on an analysis of existing literature concerning AI applications in PM from 2016 to 2023. Table 1 highlights the strengths and weaknesses of the chosen approaches. Table 2

exhibits the comparative evaluations of commonly employed classification techniques based on existing empirical evidence and theoretical studies [3, 4, 7]. The analysis reveals that across various datasets, no individual learning algorithm consistently outperforms all others.

**Table 1.** Strengths and weaknesses of each algorithm

Algorithm	Strengths	Weaknesses
RF	<ul style="list-style-type: none"> <li>• Random Forest's ensemble approach, which combines multiple decision trees, helps mitigate overfitting, resulting in better generalization to unseen data</li> <li>• It provides valuable insights into feature importance, allowing users to identify the most influential variables in the model's predictions</li> <li>• Random Forest can effectively manage missing data, ensuring that data imputation is not always necessary</li> <li>• Versatility</li> </ul>	<ul style="list-style-type: none"> <li>• The ensemble of decision trees can make Random Forest models more complex, which can be a drawback for interpretability in some cases</li> <li>• Random Forest models can consume more memory due to their ensemble structure, which can be a concern when working with limited resources</li> <li>• In datasets with significant noise or outliers, Random Forests can still be prone to overfitting, although to a lesser extent compared to individual decision trees</li> </ul>
SVM	<ul style="list-style-type: none"> <li>• Effective in High-Dimensional Spaces: SVM perform well in high-dimensional feature spaces, making them suitable for tasks with many attributes</li> <li>• SVM aim to find the global optimal solution, ensuring a high-quality decision boundary</li> <li>• SVM often provide good generalization performance, making them useful for out-of-sample predictions</li> </ul>	<ul style="list-style-type: none"> <li>• SVM can be computationally intensive and may struggle with large datasets</li> <li>• SVM require careful parameter tuning, and choosing the appropriate kernel function is often a non-trivial task</li> <li>• The decision boundaries created by SVMs can be complex and difficult to interpret, limiting their transparency in some cases</li> </ul>
ANN	<ul style="list-style-type: none"> <li>• ANN can adapt and learn from new data, allowing them to continuously improve their performance</li> <li>• ANN can capture complex non-linear relationships in data, making them versatile for a wide range of tasks</li> <li>• ANN can be parallelized efficiently, enabling faster training and inference on modern hardware</li> </ul>	<ul style="list-style-type: none"> <li>• Selecting the appropriate network architecture and hyperparameters can be challenging and time-consuming</li> <li>• Training deep ANNs can be computationally expensive and may require high-performance hardware</li> </ul>



**Table 2.** Comparative between ML algorithms

	RF [3]	SVM [4]	NN [7]
Learning speed	Very good	Good	Average
Accuracy	Good	Excellent	Very good
Robustness against missing values	Very good	Good	Good
Resilience to irrelevant attributes	Very good	Excellent	Excellent
Resilience to redundant attributes	Good	Very good	Very good
Resilience to highly interrelated attributes	Good	Very good	Very good
Handling discrete, binary, and continuous attributes	All	Not discrete	Not discrete
Resilience to noise	Good	Good	Good
Addressing the risk of overfitting	Good	Good	Average
Efforts towards incremental learning	Good	Good	Very good
Clarity and transparency of knowledge or classification	Excellent	Average	Average
Multiclassification	Excellent	Binary classifier	Naturally extended
Classification Speed	Excellent	Excellent	Excellent

## 4 Conclusion

This paper presents a multi-criteria analysis and comparative study that focuses on customer classification models in marketing. The research starts by identifying relevant studies that employ various customer classification models. The models mentioned in these studies are then presented, along with their respective strengths and weaknesses as discussed in this article. Subsequently, specific criteria are established for each model to facilitate the comparison. According to the findings of our comparison, the SVM model emerges as the top-performing model due to its exceptional accuracy and fast learning capabilities. This suggests that the SVM model outperforms other models, which may be slower and less accurate, particularly in real-time scenarios. This study provides valuable guidance for researchers and scientists, enabling them to select the most suitable models and algorithms based on their specific requirements and preferred criteria. The primary objective of this study is to assist researchers and practitioners in making informed decisions regarding the selection of the most efficient model for their works and projects.



## References

1. El Koufi, N., Belangour, A., Sadiq, M.: A systematic literature review of machine learning techniques applied to precision marketing. *Tech. Phys. Probl. Eng. (IJTPE)*. **14**(9), 104–110 (2022)

2. Zhang, M.: Research on precision marketing based on consumer portrait from the perspective of machine learning. *Wirel. Commun. Mob. Comput.* **2022**(10), 1–10 (2022)
3. Zhang, Y.: Prediction of customer propensity based on machine learning. In: *The Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS)*, vol. 21, no. 11, pp. 5–9. IEEE (2021)
4. Zhu, G., Wu, Z., Cao, J., Gu, J.: Online tourism purchase analysis and prediction. In: *The Sixth International Conference on Advanced Cloud and Big Data (CBD)*, vol. 18, no. 09, pp. 171–176. IEEE (2018)
5. El Koufi, N., Belangour, A., Sadiq, M.: Research on precision marketing based on big data analysis and machine learning: case study of Morocco. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)* **13**(10), 58–63 (2022)
6. El Koufi, N., Belangour, A.: Artificial intelligence techniques applied in precision marketing: a survey. In: *2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, vol. 23, no. 90, pp. 1–5. IEEE, July 2023
7. Jiang, L.: Support vector machine (SVM) marketing strategy analysis method based on time series. In: *IOP Conference Series: Materials Science and Engineering*, vol. 750, no. 1, pp. 1–6 (2020)
8. Zhang, S., Liao, P., Ye, H.Q., Zhou, Z.: Multiple resource allocation for precision marketing. In: *IOP Conference Series: Materials Science and Engineering*, vol. 1592, no. 1, pp. 1–15 (2020)
9. Chong, A.Y.L., Ch'ng, E., Liu, M.J., Li, B.: Predicting consumer product demands via big data: the roles of online promotional marketing and online reviews. *Int. J. Prod. Res.* **55**(17), 5142–5156 (2017)
10. Liu, S.Y.: Precision marketing scheme based on integrating spatio-temporal data clustering and neural network. *J. Phys. Conf. Ser.* **1087**(3), 1–7 (2018)



# Artificial Intelligence in Industrial Internet of Things: A Concise Review of Performance Management

Seda Balta Kaç  and Süleyman Eken  

Kocaeli University, 41001 Izmit, Kocaeli, Turkey  
suleyman.eken@kocaeli.edu.tr

**Abstract.** The success of the IoT applications in the field of Information Technologies has led to its spread to different areas of use. In this study, the performance management of infrastructures developed using Industrial IoT sensors is examined. First, the impact of performance management in different sectors is explained. Then, different definitions and types of maintenance in the literature are explained comparatively. Then, big datasets obtained with Industrial IoT sensors and their applications are discussed. Also, studies conducted in different fields with different methods using trend monitoring applications are mentioned. The common aspect of these studies is that they provide applications that increase performance management through trend monitoring. The combination of all these concepts and technologies represents a positive effect. For the sustainability of this effect and performance management, decision making strategies in the predictive maintenance of the devices are mentioned.

**Keywords:** Predictive Maintenance · Real-time Data Sensing · Industrial Control Systems · IIoT · Trend Monitoring · Remaining Useful Life

## 1 Introduction

With the concept of the Internet of Things (IoT), many devices have become a collection of interconnected systems [1]. The development of technology increases the use of IoT in different fields and sectors, from household appliances to critical infrastructures. This technology, which is mostly used to facilitate our daily activities, also provides monitoring and control services to the user. The concept of the IoT was initially focused on the user. However, this great success of IoT has led to its expansion and use in other areas such as industrial sectors [2]. The use of IoT technology along with Industry 4.0 is called Industrial Internet of Things (IIoT) [3]. IIoT is a collection of systems that bring together different components such as smart sensors, machines, tools, embedded systems, and cloud servers, just like Internet of Things. The data obtained by smart sensors is stored on cloud servers continuously and in real time. The stored big data is processed with very different artificial intelligence techniques and turned into a meaningful whole. IIoT has various uses, such as robotic systems, autonomous vehicles, smart manufacturing,

smart cities, quality management and predictive maintenance. The main objectives of industrial institutions and organizations are to achieve maximum efficiency by monitoring, controlling and managing established infrastructures. The most important factor to achieve this goal is good performance management. The emergence of the IoT expresses exactly this. Performance management is a process that includes monitoring, controlling and managing the performance of infrastructures developed in IoT. This process aims to ensure that the established infrastructures operate with maximum efficiency and effectiveness. In addition, performance management practices have a significant impact on the reliability, sustainability and availability of certain industrial processes and systems. Organizations want to create an information community by analyzing the data obtained from the created infrastructures in real time. The main aim of this study is to provide good performance management with instant and future-oriented inferences by using the information obtained. Good performance management for infrastructures brings many successes such as continuity, financial gain and time saving. Industrial IoT sensors cause different performance increase effects in every area they are used. Considering the infrastructures that appeal to different sectors, the performance increases provided by IIoT can be summarized as follows.

**Logistics:** The physical behavior of the driver can be observed instantly through the IoT sensors (ship, truck, train, etc.) placed on the vehicles carrying the inventory. Accordingly, plans can be made to prevent possible accidents. For example, Wang et al. [4] analyzed the competitiveness of fruit and vegetable trade between Pakistan and China. As a result, it has been revealed that the competitiveness between the two countries is weak. However, the commercial structure of fruit and vegetable exchange between China and Pakistan is quite similar, consistent and open to cooperation. According to the article, establishing a logistics network based on the IIoT between the two countries will increase trade in agricultural products.

**Health:** The hand and eye coordination of the inpatient can be monitored through sensors placed in the room. In this way, action can be taken according to the physical behavior of the patient. For example, babies cannot express themselves. But they have their own methods of expression. Yurtsever and Eken [5] studied infants' body movements (such as back throwing, head banging, leg kicking, eye rubbing, stretching, thumb sucking). The study is a guide for parent-infant communication.

**Critical Infrastructures:** Water resources can be monitored and managed in real-time by equipping them with various sensors. In this way, physical, biological or cyberattacks on water systems can be monitored or intervened immediately. Hemdan et al. [6] presented a study on water quality analysis using IoT and big data analytics. According to the data of the US Environmental Protection Agency, the parameters affecting the quality of the water were selected. Then, considering these parameters, an IoT based model is proposed. They suggested that with this model, the quality of the water could be predicted online. Balta et al. [7] proposed a model for instant attack detection and real-time monitoring of cyberattacks on Supervisory Control and Data Acquisition (SCADA) systems. Distributed topic-based messaging is used in the model. Data flowing instantly from SCADA systems are analyzed with deep learning and machine learning techniques. The proposed model has been tested on water distribution systems. Study performance gives better results than other studies in the literature. Khan et al. emphasize the importance

of the digital twin model in ensuring performance management of established physical systems in their study. The research presents an explanatory architecture for implementing digital twin-based applications [8]. Haseeb-Ur-Rehman et al. conducted a study to improve the performance of wireless sensor network integration with cloud applications. The research focuses on detailing and analyzing existing integration frameworks. The aim of the study is to provide guidance for the design of cloud integration frameworks [9].

**Traffic:** Driver behavior and traffic density can be examined with IoT systems installed in certain places. Online changes can be made in the lighting systems according to the traffic density. Ouallane et al. [10] emphasize that traffic management will be better provided by including cameras used in smart city applications at intersections. In this way, the causes of road congestion will be determined more easily and action will be taken more quickly. In addition, in future studies, fuel consumption and environmental pollution management for a green city can be achieved thanks to these monitoring systems.

**Automotive:** Customer satisfaction can be aimed at the high quality production of accessories used in the automotive industry. Gopalakrishnan and Kumaran [11] conducted a study using IoT sensors to achieve this positive process. In the study, operational data such as humanity, viscosity, and force were continuously collected from the machines in the production line. With these data, they made a predictive maintenance approach. They have predicted the failures that may occur in the production line.

**Wearable technology:** Recently, smart watches used to track personal activity and health information are very popular. Smart watches can also be shown as an example of IoT usage. By using industrial sensors, studies can be carried out in the field of textiles. Milovic et al. [12] have used textile pressure sensors to predict disease related to people's gait disturbances.

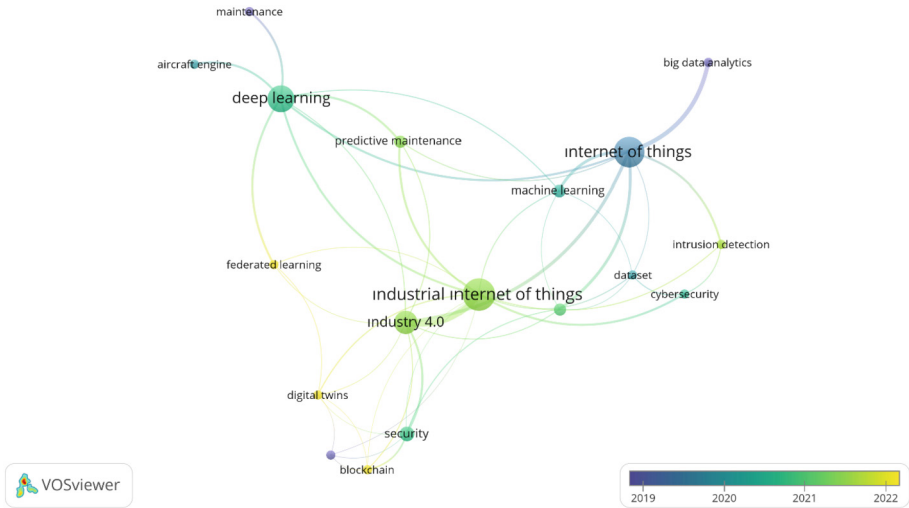
The contributions of the related study to the literature can be detailed as follows:

- Providing a comprehensive analysis of performance management in infrastructures developed using industrial IoT sensors.
- Focusing on the concept of “maintenance” as a crucial parameter in performance management.
- Enhancing performance management through the application of trend monitoring techniques.
- Analyzing the positive effects of combining these concepts and technologies on performance management.
- Exploring decision-making strategies in predictive maintenance planning to ensure the sustainability of performance management.

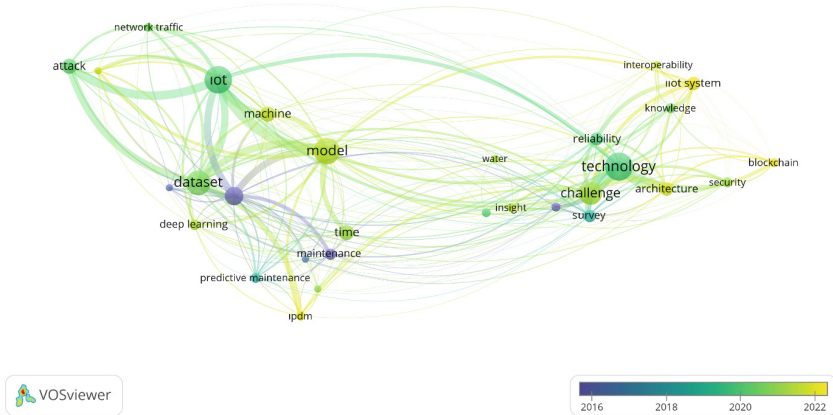
These contributions aim to contribute to the existing literature by demonstrating how performance management applications are utilized in various fields and industries for critical infrastructures. The study also references examples from the literature, including IoT, deep learning, hybrid models, and machine learning, to provide a broader perspective. The article concludes by summarizing the findings and offering insights for future research directions.

**Motivation of The Paper**

The motivation of this study is to present the positive effects of industrial internet of things technology and performance management applications together and form a guiding guide. In the literature, there some reviews/surveys in IIoT technologies and its applications. Table 1 shows the comparison with other review and survey papers. Figures 1 and 2 shows keywords and abstract analysis of the all papers in this review, respectively. They were obtained via WOS Viewer.



**Fig. 1.** Keyword analysis of all references by year



**Fig. 2.** Abstract analysis of all references by year

The remainder of this article is organized as follows. In Sect. 2, different definitions and types of maintenance in the literature are mentioned. In Sect. 3, the similarities/differences to the definition of Information Technology (IT) and Operational

Technology (OT) are mentioned. In addition, examples of large data sets obtained using IIoT infrastructures are given. In Sect. 4, examples of studies in the fields of IoT and deep learning, hybrid models and machine learning are given. In Sect. 5, decision-making strategies for the maintenance applications of devices in systems designed using IIoT are expressed. The last section concludes the article and gives direction to future work.

## 2 Maintenance Strategies for Efficient Infrastructures

Each corporate system has a physical infrastructure consisting of sensors, devices and components. Malicious biological, physical or cyberattacks can be organized on systems. These physical infrastructures should have the ability to respond quickly to sudden problems that may occur in the system. Online monitoring systems have been established in order to observe and react instantly to sudden changes that may occur in the established infrastructures.

Previously, factories had single-band production systems. In this way, it was possible to visually control the productions by people. With the increase in population and the advancement of technology, new needs and opportunities have emerged. The increase in production has revealed more manpower capacity. Automation communities consisting of many components and devices have now been formed, instead of systems controlled from a single band and from a single location. For this reason, the observed area has grown and control has become difficult. For these reasons, there was a need for systems that could remotely control the entire system online and allow immediate intervention when necessary. SCADA and Human Machine Interface (HMI) systems have been developed in response to these needs. Thanks to these systems, existing infrastructures can be monitored and observed remotely and the system can be intervened if necessary. These created systems bring many benefits to their users. The main purpose of these benefits is undoubtedly to increase productivity. In order to increase productivity, it is essential to improve the performance of the systems created [24].

Aging infrastructures face natural challenges as they near the end of their life. The possibility of failure and deterioration of aging infrastructures should be considered and maintenance measures should be taken accordingly to increase productivity. When the right precautions and precautions are not taken, losses are experienced in production, time, material and vital issues. For example, a 2018 survey study concluded that US\$472.6 billion will be needed over the next 20 years to maintain and improve drinking water infrastructure in the United States. The majority of this expenditure (US\$312.6 billion) is for the replacement or replacement of aging or deteriorated distribution assets [25]. For this reason, maintenance practices are vital for the correct, error-free and sustainable performance of an infrastructure.

Since the 1990s, different views and strategies for care management have been proposed in the literature. According to Suzuki [26], when certain activities are performed in a planned manner, they increase output (minimum failures, zero defects and zero abnormalities). Bateman [27] examines the types of care under three headings: reactive, preventive and predictive. Among these types of maintenance, malfunctions can be prevented with preventive and predictive maintenance applications.

**Table 1.** Comparison with other review and survey papers.

Paper title	Year	Focus
AI-inspired non-terrestrial networks for IIoT: Review on enabling technologies and applications [13]	2020	The focus is on the challenges that exist in the combination of complex infrastructure of non-terrestrial networks (NTNs) with IIoT systems. In this context, the contribution of artificial intelligence (AI) techniques is mentioned
A comprehensive survey on interoperability for IIoT: taxonomy, standards, and future directions [14]	2021	It focuses on solution proposals to increase the interoperability of different IIoT components
Industrial internet of things (IIoT) applications of edge and fog computing: A review and future directions [15]	2021	The impact of IIoT on edge and fog computing (security, privacy, compliance, virtualization, etc.) and applicability are discussed
Recent Technologies, Security Countermeasure and Ongoing Challenges of Industrial Internet of Things (IIoT): A Survey [16]	2021	It focuses on the vulnerabilities and security measures of the IIoT architecture. Solution approaches to ongoing security challenges are presented
A Comprehensive Survey of Digital Twins and Federated Learning for Industrial Internet of Things (IIoT), Internet of Vehicles (IoV) and Internet of Drones (IoD) [17]	2022	The study focuses on federated learning and digital twin applications for IIoT, Internet of Vehicles (IoV) and Internet of Drones (IoD)
Key Challenges and Emerging Technologies in Industrial IoT Architectures: A Review [18]	2022	3 different IIoT architectures are analyzed considering the concepts of scalability, interoperability, security, privacy, reliability and low latency. New technologies that address these challenges (edge/fog computing, blockchain, SDN, Machine Learning etc.) are mentioned
A comprehensive survey on blockchain in industrial internet of things: Motivations, research progresses, and future challenges [19]	2022	Researches covering the coexistence of Blockchain and IIoT are focused on different perspectives such as motivation, benefit and requirements
URLLC and eMBB in 5G Industrial IoT: A survey [20]	2022	This article presents a comprehensive survey of 5G powered IIoT systems
Review of commercial and open technologies available for Industrial Internet of Things [21]	2022	The study focused on the technologies, differences and challenges needed to develop a complete IIoT environment

*(continued)*



**Table 1.** (continued)

Paper title	Year	Focus
Knowledge-Based Fault Diagnosis in Industrial Internet of Things: A Survey [22]	2022	The focus is on IIoT systems designed using knowledge-based error detection approaches. Successful implementations and ongoing issues are covered
On the Reliability of Industrial Internet of Things from Systematic Perspectives: Evaluation Approaches, Challenges, and Open Issues [23]	2022	In this study, different performance improvement methods for safer IIoT systems are examined. The focus is on a single framework proposal that will ensure all aspects of security

- **Predictive Maintenance** The physical condition of the equipment under different conditions (such as temperature, erosion, noise) is taken into account. Maintenance planning is made in accordance with these situations. It is quite similar to preventive maintenance. The disadvantage of preventive maintenance is given to production at scheduled intervals to get the job done. In predictive maintenance planning, instead of waiting for a period of time, maintenance is planned for real physical conditions and needs [28].
- **Breakdown Maintenance** (BM) is performed after system performance becomes dangerous.
- **Regular preventive maintenance** practices have been diversified into time-based, usage-based, and condition-based to reduce pre-failure risk.
  - Time-Based Maintenance (TBM) includes preventive maintenance practices planned according to certain time intervals,
  - Usage-Based Maintenance (UBM) (Usage-Based Maintenance) maintenance practices planned according to equipment usage measures,
  - Condition Based Maintenance (CBM) (Condition-Based Maintenance) also covers the planned maintenance applications according to the actual faults on the equipment.
- **Corrective Maintenance** (CM) is planned maintenance practices based on continuous improvement [25].
- **Reactive maintenance** The components in the system are allowed to work until they fail. Corrective maintenance is expressed as the maintenance performed after a malfunction occurs. The corrective maintenance approach is based on seeking a solution when a malfunction occurs in the system. After failure, the relevant component is repaired or replaced. For this reason, reactive maintenance increases costs.

Swanson [29] has examined care strategies under three headings as reactive, proactive and aggressive. Unlike other studies, he defined preventive and predictive maintenance

practices as proactive care. Aggressive maintenance, on the other hand, aims to improve performance while continuing to prevent failures [30].

Unlike other studies, predictive maintenance practices enable real-time detection of malfunction and inefficiency problems, optimizing pre-failure maintenance planning and extending the life of the devices used. Malfunctions and deteriorations that may occur in the production lines of the factories may stop the production. Stopping of production not only incurs the expense of the relevant component, but also incurs additional costs arising from production. Monitoring equipment and preventing possible failures increase efficiency in infrastructures.

Using the IoT, data is received online from the sensors. These read data are tried to be made meaningful by using approaches such as artificial intelligence, machine learning, and deep learning together with the previously received data. Predictive maintenance forecasts for the devices in the system are made with this data, which is made meaningful by reading from the past data. With these estimations, the average deterioration time of the parts in the system, probability of deterioration and instant random anomaly detection are made. With these predictions, failure times for devices can be determined in advance and measures can be taken accordingly. In this way, organizations can provide time savings and cost savings with the right planning. Predictive maintenance practices also provide efficiency, savings and occupational safety advantages.

### **3 Large Datasets from Both IT and OT Systems from Field Assets**

IT encompasses a wide range of communication tools and protocols, network hardware like cables and switches, and data storage devices like servers and software. Information is typically handled for business purposes, and IT networks and devices collaborate to store, regulate, and transport information. A couple of instances of software utilized in IT include material requirements planning (MRP) and enterprise resource planning (ERP) systems. These kinds of software operate at the macro level to handle strategies, calculations, and other high-level operations used to make commercial decisions or oversee industrial processes. These systems require a lot of hardware, thus maintaining them costs money. Moreover, rather than assessing the material, they emphasized providing a “roadway” for managing it.

IT has played a significant role in the development of industrial enterprises for many years. With it, businesses have enhanced procedures and introduced automation that drew on the most cutting-edge technologies. However, technology is advancing, and operational technology, sometimes known as OT, has overtaken IT systems in terms of the extensibility of data from the machines and equipment that power industrial control systems. The capacity to gather data and manage physical devices on industrial equipment, when combined with the IIoT, has transformed the idea of production and enabled deep insight across activities.

The way data is used is the main distinction between IT and OT. IT is more concentrated on general business requirements. It handles voice communication, data storage (typically in unstructured databases), transactions, and other meta-level data requirements. OT, in contrast, deals with data generated by machines and intended for immediate consumption by users or managers. This information is derived through the control

of physical objects by means of digital technologies, such as software equipped with highly developed analytics engines for process optimization [31–33].

Alsaedi et al. [34] offer a current and representative dataset that may be utilized to precisely plan and assess IoT/IIoT defense systems. For each IoT device, they suggest a new data-driven IIoT-based dataset (telemetry data) that was acquired from a sample scale-down testbed. Seven IoT and IIoT sensors, including weather and Modbus sensors, were employed in the testbed to collect their telemetry data. Meidan et al. [35] built a dataset to test a network-based anomaly detection method that retrieves behavioral snapshots from the network and uses deep autoencoders to identify unusual network traffic coming from hacked IoT devices. Nine IoT devices made up the lab environment that was developed. In order to find and identify botnets on IoT-specific networks, Koroniotis et al. [36] created a dataset utilizing both real and simulated IoT network traffic. Three components make up the lab environment. MQTTset is an IoT dataset created by Vaccari et al. [37] that focuses on Message Queuing Telemetry Transport (MQTT) communications. IoT-Flock [38], a network traffic generation program that can simulate IoT devices and networks based on the MQTT and CoAP protocols, was used to build MQTTset. Al-Hawawreh et al. [39] created a unique IIoT intrusion data set that captures the alterations and heterogeneity of system and network activity brought on by diverse IIoT devices, connectivity protocols, and communication patterns. They also provided a number of attack scenarios and fresh assaults pertaining to IIoT connectivity standards. Ferrag et al. [40] created a remarkably original IoT and IIoT cybersecurity dataset that accurately captures the key IoT characteristics and diverse network traffic. The many IoT devices produce the IoT data (more than 10 types). Besides existing datasets related to cyberattacks, there are different data sources [41–44] such as motor test-rig, armored vehicle, Computer Numerical Control (CNC) machine, locomotive bearing, and wind turbine for predictive maintenance based on data-driven methods.

## 4 AI-Enabled Approaches for Trend Monitoring

### 4.1 Trend Monitoring

Trend monitoring is especially used in operational technology systems. It provides online monitoring of critical infrastructures installed by using different devices such as sensors, operators, controllers for a specific purpose. Thanks to trend monitoring, real-time data obtained through sensors can be monitored instantly and the system can be intervened when necessary. At the same time, the data obtained through trend tracking applications are combined with techniques such as artificial intelligence, big data analytics, machine learning and deep learning to make them meaningful.

Since the data obtained from most infrastructures is real-time and streaming data, it can be said that it is quite large in volume. It seems very difficult to make sense of this data, which is growing day by day, by scanning it with a human eye. For these reasons, the data obtained from the system can be converted into information by using algorithms and techniques related to these reasons. Data is considered as an unprocessed raw concept. Drawing meaningful conclusions from this data can be defined as knowledge. These meaningful results, obtained using trend monitoring and Artificial Intelligence

technologies, enable future predictions about anomaly detection, maintenance planning and the status of systems.

## 4.2 Traditional Machine Learning Methods

In this section, examples of studies using trend monitoring and traditional machine learning methods and their contributions are given. Qaiser et al. focus on examining DDoS vulnerabilities that can adversely impact Industrial Internet of Services (IioS) techniques. The study involves classifying data traffic using various machine learning algorithms [42]. Albalzahr et al. propose an IoT-based monitoring and early warning system (EEWS) to protect human lives from earthquake disasters. The activities of the monitored area are tracked online through IoT sensors. The observed parameters are then analyzed using linear and non-linear machine learning algorithms to derive meaningful insights [45].

Arowolo et al. developed a machine learning application using the San Francisco COVID-19 dataset, which contains information obtained by IoT systems during the Covid-19 period. The study focuses on the use of the Artificial Bee Colony (ABC) algorithm for dimensionality reduction and the Support Vector Machine (SVM) algorithm for classification [46]. Tekin et al. proposed a machine learning-based intrusion detection system to address security and privacy concerns in smart home systems (SHSs). The study also analyzes the energy consumption of the proposed system using cloud computing-based, IoT device-based, and edge computing-based approaches. Considering all the constraints, the Decision Tree (DT) algorithm provides the best results [47, 48].

## 4.3 Deep Learning Methods

In this section, examples of studies using trend monitoring and Deep Learning methods and their contributions are given. Abusitta et al. perform anomaly detection on data obtained from systems designed using IoT technology. They employ the denoising autoencoder model to extract features from real-world datasets in noisy environments. Subsequently, a classifier is used to distinguish the anomaly status of the data [46]. Kolosnjaji et al. model the underlying patterns of malware under large datasets using CNN and RNN deep learning methods to detect malicious software [49].

Saranya et al. discuss the possibility of achieving Precision farming by combining deep learning and IoT applications. They compare different datasets used in smart agriculture. They utilize the VGG16 transfer learning model to achieve high accuracy in plant pest detection [50]. Naseri et al. propose a deep learning and IoT-based face mask detection model to protect human health and prevent the spread of viruses. They optimize various parameters in Single Shot Multi-box Detector (SSD), ResNet, and MobileNet using a method called Adaptive Sailfish Moth Flame Optimization (ASMFO). The proposed model is analyzed on three different datasets [51]. Rajkumar et al. utilize the Hungarian heart disease dataset collected from IoT devices to predict heart disease. They employ the Median Studentized Residual method to handle erroneous and missing data, the Harris Hawk Optimization (HHO) approach for feature selection, and the Modified Deep Long Short-Term Memory (MDLSTM) for classification [52].

#### 4.4 Hybrid Methods

In this section, examples of studies using trend monitoring and hybrid methods and their contributions are given. Yazdinejad et al. propose an ensemble-based deep learning model that combines Long Short-Term Memory (LSTM) and AutoEncoder (AE) architectures to detect anomalous activities during cyber attacks in IIoT. LSTM is used for learning on time series data, while AutoEncoder is employed for feature selection. The proposed model is evaluated on real IIoT datasets and achieves superior results compared to other machine learning algorithms [53].

Jahromi et al. propose a model for detecting cyber attack types. The proposed model, called deep federated learning cyber-threat hunting model, consists of two main components representing normal and abnormal states. The proposed model has been observed to have faster training time compared to other existing studies when compared [54]. Sankaran and Kim propose a method for anomaly detection and secure data transmission in Industrial IoT applications. They introduce the Robust Multi-cascaded CNN (RMC-CNN) approach to detect attacks against the identified network and determine the type of attack. The obtained data is then encrypted using the dynamic honey pot encryption algorithm to ensure secure transmission [55].

### 5 Decision Making Strategies

- This section explains the decision making strategies. Systems developed for OT applications include quite a lot of equipment and components. Often the prices of these equipments are quite high. For these reasons, organizations aim to extend the life of the equipment as much as possible instead of replacing the equipment in the infrastructure. Performance information is often needed to plan and control a maintenance process. Different methods can be tried to obtain the maintenance performance information of the devices. In the new technological era, IoT, and IIoT applications have become quite popular, as given above. The data obtained through these applications is very valuable for generating maintenance performance information.
- Predictive planning can be made with the relevant metrics from the data obtained from the infrastructure created first. However, collecting data from a system is a laborious and lengthy process. It also requires experience. In order to obtain performance information, the systems installed need a certain wear time. In this context, maintenance practices for newly installed systems become critical. Maintenance of equipment in a system to be designed for the first time can be included in the predictive maintenance approach.
- Compare et al. [56] conducted a study describing the challenges, strengths, advantages and disadvantages of the predictable maintenance paradigm in 2020. Various concepts are discussed for the successful implementation and dissemination of predictive maintenance (PdM) in applications involving the IoT. There are basically 4 different stages for PdM. It is emphasized that the right components should be selected first for a good PdM. After choosing the right components, it is necessary to determine the most suitable critical infrastructures. After the appropriate infrastructures are prepared, the data obtained with IoT should be combined with the correct algorithms and analyzes. Finally, the contribution of real-time IoT-enabled remote monitoring systems is mentioned after all the steps for good predictive maintenance have been performed.

## 6 Conclusions

In this study, firstly, the importance of performance management is explained based on different perspectives. Maintenance practices are very important for good performance management. There are many different definitions and classification information in the literature on maintenance performance. The types of care in the literature are explained in detail.

It is anticipated that the proposed method will enable predictable maintenance planning for better performance management in Industrial IoT applications. The similarities/differences between these concepts and their effects on each other are discussed. Then, sample large data sets obtained using Industrial IoT technology are introduced. The effect that occurs when industrial IoT sensors and trend monitoring approaches are used together is exemplified by various literature reviews. As can be seen from the related studies, the coexistence of these two has very positive effects.

The combined application of IIoT and trend monitoring applications makes it possible to control the infrastructure of physical systems, monitor them remotely and intervene in the system when necessary. The realized systems consist of different components (motor, sensor, unit, etc.) and equipment. There are many devices in the infrastructures of large organizations. Performance management of these devices can be achieved with predictable maintenance practices. The best plans for predictable maintenance approaches are provided by data obtained from infrastructures. Data is analyzed using artificial intelligence, machine learning, deep learning or hybrid models. Making a predictive maintenance plan as a result of the analyzes improves the performance of the systems. However, on newly installed systems, predictable maintenance planning will be challenging when data collection is time consuming. At this stage, new data sets covering different conditions can be created by making use of previous large data sets and expert knowledge. When the systems are detailed and well adopted, the synthetic datasets will be likely to be close to reality.

**Acknowledgments.** This study was supported by The Scientific and Technological Research Council of Turkey (TUBITAK) with project number 122E610.

## References

1. Atzori, L., Iera, A., Morabito, G.: The internet of things: a survey. *Comput. Netw.* **54**(15), 2787–2805 (2010). <https://doi.org/10.1016/j.comnet.2010.05.010>
2. Serror, M., Hack, S., Henze, M., Schuba, M., Wehrle, K.: Challenges and opportunities in securing the industrial internet of things. *IEEE Trans. Ind. Inf.* **17**(5), 2985–2996 (2020)
3. Xu, L.D., He, W., Li, S.: Internet of things in industries: a survey. *IEEE Trans. Ind. Inf.* **10**(4), 2233–2243 (2014)
4. Wang, B., Zhang, Y.T., Wang, T., Li, H., Zhou, C. M., et al.: Analysis of competitiveness and complementarity of Chinese fruits and vegetables in Pakistani market in the context of industrial internet of things. *Mob. Inf. Syst.* (2022). <https://doi.org/10.1155/2022/3189858>
5. Yurtsever, M.M.E., Eken, S.: BabyPose: real-time decoding of baby's non-verbal communication using 2D video-based pose estimation. *IEEE Sens. J.* **22**(14), 13776–13784 (2022). <https://doi.org/10.1109/JSEN.2022.3183502>

6. Hemdan, E.E.D., Essa, Y.M., El-Sayed, A., Shouman, M., Moustafa, A.N.: Smart water quality analysis using IoT and big data analytics: a review. In: Proceedings of the 2021 International Conference on Electronic Engineering (ICEEM), Menouf, Egypt, pp. 1–5 (2021). <https://doi.org/10.1109/ICEEM52022.2021.9480628>
7. Balta, S., Zavrak, S., Eken, S.: Real-time monitoring and scalable messaging of SCADA networks data: a case study on cyber-physical attack detection in water distribution system. In: Proceedings of the International Congress of Electrical and Computer Engineering, Bandırma, Turkey, pp. 203–215 (2022). [https://doi.org/10.1007/978-3-031-01984-5\\_17](https://doi.org/10.1007/978-3-031-01984-5_17)
8. Khan, L.U., et al.: Federated learning for digital twin-based vehicular networks. Architecture and challenges. *IEEE Wirel. Commun.* (2023). <https://doi.org/10.1109/MWC.012.2200373>
9. Haseeb-Ur-Rehman, R.M.A., et al.: Sensor cloud frameworks: state-of-the-art, taxonomy, and research issues. *IEEE Sens. J.* **21**(20), 22347–22370 (2021). <https://doi.org/10.1109/JSEN.2021.3090967>
10. Ouallane, A.A., Bahnasse, A., Bakali, A., Talea, M.: Overview of road traffic management solutions based on IoT and AI. *Procedia Comput. Sci.* **198**, 518–523 (2022). <https://doi.org/10.1016/j.procs.2021.12.279>
11. Gopalakrishnan, S., Kumaran, M.S.: IIoT framework based ML model to improve automobile industry product. *Intell. Autom. Soft Comput.* **31**(3), 1435–1449 (2022). <https://doi.org/10.32604/iasc.2022.020660>
12. Milovic, M., Fariñas, G., Fingerhuth, S., Pizarro, F., Hermosilla, G., et al.: Detection of human gait phases using textile pressure sensors: a low cost and pervasive approach. *Sensors* **22**(8), 2825 (2022). <https://doi.org/10.3390/s22082825>
13. Michailidis, E.T., Potirakis, S.M., Kanatas, A.G.: AI-inspired non-terrestrial networks for IIoT: review on enabling technologies and applications. *IoT* **1**(1), 3 (2020). <https://doi.org/10.3390/iot1010003>
14. Hazra, A., Adhikari, M., Amgoth, T., Srirama, S.N.: A comprehensive survey on interoperability for IIoT: taxonomy, standards, and future directions. *ACM Comput. Surv. (CSUR)* **55**(1), 1–35 (2021). <https://doi.org/10.1145/3485130>
15. Chalapathi, G.S.S., Chamola, V., Vaish, A., Buyya, R.: Industrial internet of things (IIoT) applications of edge and fog computing: a review and future directions. In: *Fog/Edge Computing for Security, Privacy, and Applications*, pp. 293–325 (2021). [https://doi.org/10.1007/978-3-030-57328-7\\_12](https://doi.org/10.1007/978-3-030-57328-7_12)
16. Tan, S.F., Samsudin, A.: Recent technologies, security countermeasure and ongoing challenges of industrial internet of things (IIoT): a survey. *Sensors* **21**(19), 6647 (2021). <https://doi.org/10.3390/s21196647>
17. Jamil, S., Rahman, M.: A comprehensive survey of digital twins and federated learning for industrial internet of things (IIoT), internet of vehicles (IoV) and internet of drones (IoD). *Appl. Syst. Innov.* **5**(3), 56 (2022). <https://doi.org/10.3390/asi5030056>
18. Mirani, A.A., Velasco-Hernandez, G., Awasthi, A., Walsh, J.: Key challenges and emerging technologies in industrial IoT architectures: a review. *Sensors* **22**(15), 5836 (2022). <https://doi.org/10.3390/s22155836>
19. Huo, R., Zeng, S., Wang, Z., Shang, J., Chen, W., et al.: A comprehensive survey on blockchain in industrial internet of things: Motivations, research progresses, and future challenges. *IEEE Commun. Surv. Tutor.* **24** (2022). <https://doi.org/10.1109/COMST.2022.3141490>
20. Khan, B.S., Jangsher, S., Ahmed, A., Al-Dweik, A.: URLLC and eMBB in 5G industrial IoT: a survey. *IEEE Open J. Commun. Soc.* **3**, 1134–1163 (2022). <https://doi.org/10.1109/OJCS.2022.3189013>
21. Schuh, G., Jarke, M., Gützlaff, A., Koren, I., Janke, T., Neumann, H.: Review of commercial and open technologies available for Industrial Internet of Things. In: *Design and Operation of Production Networks for Mass Personalization in the Era of Cloud Technology*. Elsevier, United States, pp. 209–241 (2022). <https://doi.org/10.1016/B978-0-12-823657-4.00005-1>

22. Chi, Y., Dong, Y., Wang, J., Yu, F.R., Leung, V.C.: Knowledge-based fault diagnosis in industrial internet of things: a survey. *IEEE Internet Things J.* **9**, 12886–12900 (2022). <https://doi.org/10.1109/JIOT.2022.3163606>
23. Kim, D.S., Hoa, T.D., Thien, H.T.: On the reliability of industrial internet of things from systematic perspectives: evaluation approaches, challenges, and open issues. *IETE Tech. Rev.* **39**, 1277–1308 (2022). <https://doi.org/10.1080/02564602.2022.2028586>
24. Sharma, A.K., Shudhanshu, A.B.: Manufacturing performance and evolution of TPM. *Int. J. Eng. Sci. Technol.* **4**(3), 854–866 (2012)
25. Predictive Maintenance of Physical Assets Use Case. <https://cdn.gihub.org/umbraco/media/3184/8-predictive-maintenance-of-physical-assets-use-case.pdf>. Accessed 4 July 2023
26. Suzuki, T.: *TPM in Process Industries*. Routledge, Newyork. Accessed 22 Apr 2023 <https://www.routledge.com/TPM-in-Process-Industries/Suzuki/p/book/9781563270369>
27. Bateman, J.F.: Preventive maintenance: stand alone manufacturing compared with cellular manufacturing. *Ind. Manag. Chicago Atlanta* **37**, 19 (1995)
28. Eade, R.: The importance of predictive maintenance. *New Steel* **13**(9), 68–72 (1997)
29. Swanson, L.: Linking maintenance strategies to performance. *Int. J. Prod. Econ.* **70**(3), 237–244 (2001)
30. Weil, N.A.: Make the most of maintenance. *Manuf. Eng.* **120**(5) (1998)
31. Kamal, S.Z., Al Mubarak, S.M., Scodova, B.D., Naik, P., Flichy, P. et al.: “IT and OT convergence-Opportunities and challenges. In: *Proceedings of the SPE Intelligent Energy International Conference and Exhibition, Scotland, UK, 2016*
32. Paes, R., Mazur, D.C., Venne, B.K., Ostrzenski, J.: A guide to securing industrial control networks: integrating IT and OT systems. *IEEE Ind. Appl. Mag.* **26**(2), 47–53 (2019). <https://doi.org/10.2118/181087-MS>
33. Bécue, A., Praça I., Gama, J.: Artificial intelligence, cyber-threats and industry 4.0: challenges and opportunities. *Artif. Intell. Rev.* **54**(5), 3849–3886 (2021). <https://doi.org/10.1007/s10462-020-09942-2>
34. Alsaedi, A., Moustafa, N., Tari, Z., Mahmood A., Anwar, A.: TON\_IoT telemetry dataset: a new generation dataset of IoT and IIoT for data-driven intrusion detection systems. *IEEE Access* **8**, 165130–165150 (2020). <https://doi.org/10.1109/ACCESS.2020.3022862>
35. Meidan, Y., et al.: N-baiot—network-based detection of IoT botnet attacks using deep autoencoders. *IEEE Pervasive Comput.* **17**(3), 12–22 (2018). <https://doi.org/10.1109/MPRV.2018.003367731>
36. Koroniotis, N., Moustafa, N., Sitnikova E., Turnbull, B.: Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: bot-IoT dataset. *Futur. Gener. Comput. Syst.* **100**, 779–796 (2019). <https://doi.org/10.1016/j.future.2019.05.041>
37. Vaccari, I., Chiola, G., Aiello, M., Mongelli M., Cambiaso, E.: MQTTset, a new dataset for machine learning techniques on MQTT. *Sensors* **20**(22), 6578 (2020). <https://doi.org/10.3390/s20226578>
38. Ghazanfar, S., Hussain, F., Rehman, A.U., Fayyaz, U.U., Shahzad F., Shah, G.A.: Iot-flock: an open-source framework for IoT traffic generation. In: *2020 International Conference on Emerging Trends in Smart Technologies (ICETST), Piscataway, New Jersey*, pp. 1–6 (2020). <https://doi.org/10.1109/ICETST49965.2020.9080732>
39. Al-Hawawreh, M., Sitnikova E., Aboutorab, N.: X-IIoTID: a connectivity-agnostic and device-agnostic intrusion data set for industrial Internet of Things. *IEEE Internet Things J.* **9**(5), 3962–3977 (2021). <https://doi.org/10.1109/JIOT.2021.3102056>
40. Ferrag, M.A., Friha, O., Hamouda, D., Maglaras, L., Janicke, H.: Edge-IIoTset: a new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning. *IEEE Access* **10**, 40281–40306. (2022). <https://doi.org/10.1109/ACCESS.2022.3165809>



41. Wang, H., Zhang, W., Yang D., Xiang, Y.: Deep-learning-enabled predictive maintenance in industrial internet of things: methods, applications, and challenges. *IEEE Syst. J.* (2022). <https://doi.org/10.1109/JSYST.2022.3193200>
42. Zhang, W., Yang, D., Xu, Y., Huang, X., Zhang J., Giglung, M.: DeepHealth: a self-attention based method for instant intelligent predictive maintenance in industrial Internet of Things. *IEEE Trans. Ind. Inform.* **17**(8), 5461–5473 (2020). <https://doi.org/10.1109/TII.2020.3029551>
43. Teoh, Y.K., Gill, S.S., Parlikad, A.K.: IoT and fog computing based predictive maintenance model for effective asset management in industry 4.0 using machine learning. *IEEE Internet of Things J.* (2021) <https://doi.org/10.1109/JIOT.2021.3050441>
44. Qaiser, G., Chandrasekaran, S., Chai, R., Zheng, J.: Classifying DDoS attack in industrial internet of services using machine learning. In: 15th International Conference on Computer and Automation Engineering (ICCAE), pp. 546–550 (2023)
45. Abdalzaher, M.S., Elsayed, H.A., Fouda, M.M., Salim, M.M.: Employing machine learning and IoT for earthquake early warning system in smart cities. *Energies* **16**(1), 495 (2023). <https://doi.org/10.3390/en16010495>
46. Arowolo, M.O., Ogundokun, R.O., Misra, S., et al.: Machine learning-based IoT system for COVID-19 epidemics. *Computing* **105**, 831–847 (2023). <https://doi.org/10.1007/s00607-022-01057-6>
47. Tekin, N., Acar, A., Aris, A., Uluagac, A.S., Gungor, V.C.: Energy consumption of on-device machine learning models for IoT intrusion detection. *Internet Things* **21**, 100670 (2023). <https://doi.org/10.1016/j.iot.2022.100670>
48. Abusitta, A., de Carvalho, G.H., Wahab, O.A., Halabi, T., Fung, B.C.M., Al Mamoori, S.: Deep learning-enabled anomaly detection for IoT systems. *Internet Things* **21**, 100656 (2023). <https://doi.org/10.1016/j.iot.2022.100656>
49. Kolosnjaji, B., Zarras, A., Webster, G., Eckert, C.: Deep learning for classification of malware system call sequences. In: Kang, B.H., Bai, Q. (eds.) *AI 2016: Advances in Artificial Intelligence*. AI 2016. LNCS, vol. 9992, pp. 137–149. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-50127-7\\_11](https://doi.org/10.1007/978-3-319-50127-7_11)
50. Saranya, T., Deisy, C., Sridevi, S., Anbananthen, K.S.M.: A comparative study of deep learning and Internet of Things for precision agriculture. *Eng. Appl. Artif. Intell.* **122**, 106034 (2023). <https://doi.org/10.1016/j.engappai.2023.106034>
51. Naseri, R.A.S., Kurnaz, A., Farhan, H.M.: Optimized face detector-based intelligent face mask detection model in IoT using deep learning approach. *Appl. Soft Comput.* **134**, 109933 (2023). <https://doi.org/10.1016/j.asoc.2022.109933>
52. Rajkumar, G., Devi, T.G., Srinivasan, A.: Heart disease prediction using IoT based framework and improved deep learning approach: medical application. *Med. Eng. Phys.* **111**, 103937 (2023). <https://doi.org/10.1016/j.medengphy.2022.103937>
53. Yazdinejad, A., Kazemi, M., Parizi, R.M., Dehghantanha, A., Karimipour, H.: An ensemble deep learning model for cyber threat hunting in industrial internet of things. *Digit. Commun. Netw.* **9**(1), 101–110 (2023). <https://doi.org/10.1016/j.dcan.2022.09.008>
54. Jahromi, A.N., Karimipour, H., Dehghantanha, A.: An ensemble deep federated learning cyber-threat hunting model for Industrial Internet of Things. *Comput. Commun.* **198**, 108–116 (2023). <https://doi.org/10.1016/j.comcom.2022.11.009>
55. Sankaran, K.S., Kim, B.H.: Deep learning based energy efficient optimal RMC-CNN model for secured data transmission and anomaly detection in industrial IoT. *Sustain. Energy Technol. Assess.* **56**, 102983 (2023). <https://doi.org/10.1016/j.seta.2022.102983>
56. Lacaille, J. Rabenoro, T.: A trend monitoring diagnostic algorithm for automatic pre-identification of turbofan engines anomaly. In: *Proceedings of the Prognostics and System Health Management Conference*, Chongqing, China, pp. 819–823 (2018). <https://doi.org/10.1109/PHM-Chongqing.2018.00146>



# Comparison of LDA, NMF and BERTopic Topic Modeling Techniques on Amazon Product Review Dataset: A Case Study

Salih Can Turan<sup>1</sup> , Kazım Yıldız<sup>2</sup> , and Büşra Büyüktanır<sup>2</sup> 

<sup>1</sup> Artiwise Software Technologies, Istanbul, Turkey  
can.turan@artiwise.com

<sup>2</sup> Computer Engineering, Faculty of Technology, Marmara University, Istanbul, Turkey  
{kazim.yildiz,busra.buyuktanir}@marmara.edu.tr

**Abstract.** With the developing technology, the e-commerce market is growing day by day. As of 2022, it is estimated that 19.7% of the sales in the world are made over the internet. However, there are negative elements that distinguish online sales from regular sales. Communication between the seller and the customer is more difficult on the online platform. Likewise, problems such as quality or cargo are constantly written under the product reviews. For this reason, the seller must constantly monitor customer feedback and take the necessary action. With topic modeling algorithms, user complaints can be grouped and read in groups. In this study, LDA (Latent Dirichlet allocation), NMF (Non-Negative Matrix Factorization) and BERTopic algorithms tested on Amazon product review dataset were compared. According to the results obtained, all 3 algorithms are successful and useful. The BERTopic algorithm produced more meaningful results than other algorithms according to the consistency calculation metric.

**Keywords:** LDA · NMF · BERTopic · Topic modeling

## 1 Introduction

Topic modeling is a method used to reveal topics within documents. It is a useful technique for analyzing large amounts of text data such as news articles, customer reviews, and social media posts. The purpose of topic modeling is to identify the most important words and phrases that occur together in a given set of documents and group them into separate topics. These topics can then be used to better understand the sub-themes or topics of the text data and to explore the relationships between different information. This article will provide an overview of the topic modeling process and discuss its applications and challenges. In this study, three topic modeling algorithms were applied on the Amazon product review dataset [1], and the results were examined and compared. These algorithms are LDA (Latent Dirichlet allocation) [8], NMF (Non-negative Matrix Factorization) [9–11] and BERTopic. At the end of the study, our goal is to identify negative customer review topics and what issues customers are complaining about. In

this way, large sellers will now be able to read product reviews as a group and maybe forward the complaint to the relevant departments. In addition, a classifier can be trained by assigning the topics obtained in this study to the documents. Later, when a new comment arrives, this classifier can immediately detect what the negative comment is about. When it is determined which topic the review belongs to, the seller can take quick action, analyze the problem that the customer is not satisfied with, and prevent financial loss by ensuring customer satisfaction. Due to this importance, it is very important that the reliability of the classification process to be performed is high. The results of the topic model to be used are of great importance in this respect.

## 2 Related Work

Many studies have been done in the field of topic modeling before. The history of LDA and NMF algorithms dates back to the 90s and early 2000s. This literature review also includes the much newer BERTopic algorithm, which strengthens topic modeling using transformers embeddings. Guitierrez (2020) performed topic modeling and sentiment classification on the Amazon dataset. In the study, worked separately on both negative and positive data, and the keywords of the subject were extracted by counting n-gram frequencies for topic modeling [2]. In another study, Egger et al. (2020) tested and compared LDA, NMF, Top2Vec, and BERTopic topic modeling algorithms using twitter data, and saw that BERTopic and NMF algorithms gave relatively better results. Also, an important result of Egger is that NMF revolves around its low capability to identify embedded meanings within a corpus [3]. Koruyan [4] implemented the BERTopic algorithm on user complaints using the data of a complaint platform. As a result of the study, the subject tree was shaped around cargo, order cancellation and general product features. Sangaraju [6] examined traditional techniques for topic modeling in the data he collected in the field of finance and compared it with the recently popular pre-trained models such as BERT, FinBERT, RoBERTa and DistilBERT. Since it can be used easily because there is no need for preprocess operations such as lemmatization in BERT-based models, it reveals that its implementation is simpler and its performance is stronger. In addition, the study also revealed that FinBERT trained in finance outperformed the success of models trained in the field. Abuzayed et al. [7] compared LDA, NMF algorithms, which are topic modeling algorithms on Arabic texts, and BERTopic, which uses a pre-trained language model. Normalized pointwise mutual information MPMI was used to evaluate the results.

## 3 Material and Method

In this section the dataset and methodology of the paper is given. The dataset consists of Amazon product reviews. LDA, NMF and BERTopic were implemented for topic modeling of the dataset. LDA and NMF topic modeling methods aim to divide the content of documents in the dataset into topics. BERTopic, on the other hand, performs topic modeling using the BERT (Bidirectional Encoder Representations from Transformers) language model.

### 3.1 Dataset

The fact that the data source is quite large and contains a variety of categories played an important role in the choice of this data. The Amazon dataset contains more than 5 million product reviews and each review takes up eight rows [1]. In the study, two of these eight columns were used. These columns are review and rating columns. Rating column is used for negative comments and review column is used for user comments. Since the data is not in table format but in.txt format with columns distributed in rows, a preprocessing is required first. In this process, the data source was transferred to a data frame object. For this process, the Pandas library was preferred due to its advantages such as the ability to easily shape the data, easy processing of null values, easy updating and processing of columns [12] (Table 1).

**Table 1.** Amazon dataset rows.

rating	The actual star- level rating of the item
product_id	The product ID
helpfulness	The helpfulness of the review
ID	The reviewer's ID
review_by	The reviewer's name
title	The title of the review
review_time	The date of the review
review	The review text

### 3.2 LDA

LDA aims to group documents in a dataset according to specific topics. This method identifies topics using the vocabularies of documents. LDA views the content of a document as a mixture of vocabularies that may belong to a particular topic and assigns the document to a topic using the relationships between these vocabularies. LDA is a useful tool for topic modeling because it requires no prior knowledge in vocabulary identification and can detect meaningful relationships between documents in the dataset [8].

### 3.3 NMF

Non-Negative Matrix Factorization (NMF) decomposes the data matrix  $V$  into non-negative factor matrices  $W$  and  $H$ . These factorized matrices extract meaningful information and provide an approximate factorization. The factor  $W$  represents the data matrix, while  $H$  contains the sampling of the data matrix. Their product forms the factorization of the data matrix. The non-negativity constraint ensures that the data matrix

contains only positive values, which does not allow subtractions. In this way, meaningful information is extracted from the data matrix. Figure 1 shows the working principle of the NMF model. Non-Negative (NMF, for decomposing matrices containing zero or higher values) The NMF algorithm tries to maximize the correspondence between the data, but there may be more than one solution [13, 14]. This can lead to differences between solutions and uncertainty about which solution should be used.

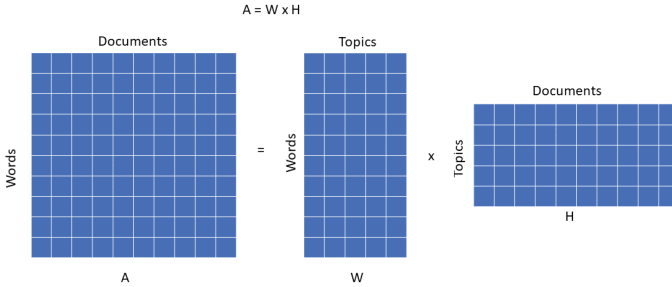


Fig. 1. NMF Working principle [18].

### 3.4 BERTopic

Figure 2 shows the BERTopic library for topic modelling as follows; BERT (Bidirectional Embeddings Representation of Transformers) input is dimension reduced with the help of UMAP (Uniform Manifold Approximation and Projection) and HDBSCAN (Hierarchical Density-Based Spatial Clustering) algorithms, and then TF-IDF and MMR are used to output the topic list [5]. Since BERTopic uses transformer embeddings, it does not need any extra cleaning in the data. However, for LDA and NMF, operations such as cleaning the punctuation marks in the data and converting the texts to lowercase letters are required. In order to avoid any problems that may arise from this situation when comparing the models, the study was carried out with the cleaned data set by performing a series of operations to clean the punctuation marks.

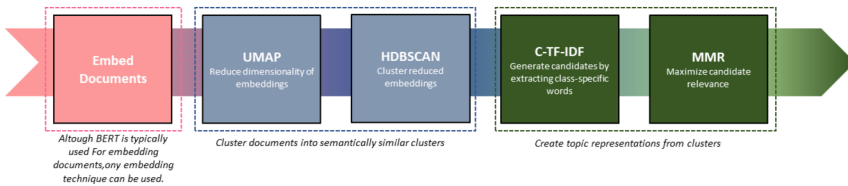
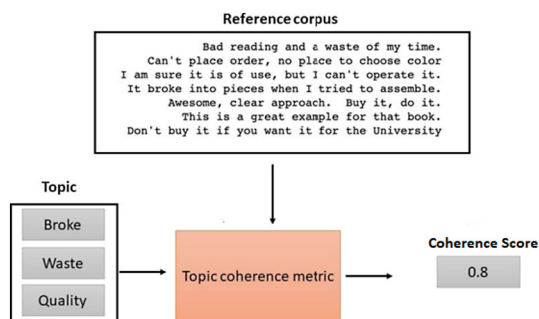


Fig. 2. BERTopic modelling architecture [15].

### 3.5 Coherence Score

The study aims to focus on product quality complaints, but excessively long comments can reduce the success. To prevent this, the success of topic modeling was increased

by applying filters to the comment length. However, the improvement rate varied across models due to differences in their structures. The main reason for this is that the BERTopic model has a filter system itself. Since the BERTopic model uses sentence transformers, it directly converts the sentence into sentence embeddings without the need for data cleaning, so it does not need data cleaning [17]. While evaluating the topic modeling, there are no direct metrics such as accuracy or f1 score like supervised learning models. In this study, we used an evaluation metric called the coherence score, but it was more accurate to make an observational evaluation in topic modeling and the results were verified by humans. A coherence score can be used as metric to measure how successful the study. Figure 3 shows the principle of the coherence score. In this case, topics are selected as the first N words that are most likely to belong to that topic. The consistency score is a measure of how similar these words are. The calculated score is not precise and may vary depending on the data being analyzed [16].



**Fig. 3.** Coherence score working principle.

The three algorithms were asked to generate a certain number of topics on the same data. These numbers are 30 75 125 and 175 respectively. Each test was performed three times and the average was taken for the coherence score. Since the BERTopic algorithm did not produce more than 175 topics, no reduction could be made, so no evaluation was made in the topic count 175 parameters.

## 4 Results and Discussion

LDA, NMF, and BERTopic algorithms were implemented on negative product reviews to capture complaint-based topics. Topic modeling was performed on reviews below a certain length (100 characters) to focus on issues like smell, fragility, and value. Comparisons showed LDA and NMF emphasized complaints, while BERTopic highlighted products. Reducing long comment inputs helped concentrate topics on complaints, as lengthy comments often focused on product features and technical details. After filtering out long comments, the topics centered on issues like fakes, malfunctions, odors, and noise. Once the model is satisfactory, it can assign subject tags to new comments for appropriate action. BERTopic also offers useful visualization tools to display hierarchical topic breakdowns.

Table 2 shows the evaluation of the methods as coherence calculation. Results were obtained by repeating and averaging 3 times for each experiment.

**Table 2.** Result of Coherence Score.

Method	Topic Count			
	30	75	125	175
LDA	0.0732	0.3011	0.3120	0.2907
NMF	0.1965	0.2236	0.2136	0.2036
BERTopic	0.5085	0.5130	0.4804	-

Since BERTopic uses sentence embeddings, it is more successful in attributing meaning to topics. The LDA algorithm has a low coherence score because it contains too many stop words in its results when the number of topics decreases. Table 3 shows the first 15 topic outputs of the LDA with topic size 30.

**Table 3.** LDA Results in 30 Titles.

Topic	Topic keywords
0	['it', 'i', 'the', 'wish', 'a', 'wish i', 'i wish', 'not', 'i wish i', 'to']
1	['it', 'i', 'the', 'not', 'to', 'this', 'a', 'and', 'my', 'is']
2	['it', 'i', 'to', 'the', 'a', 'not', 'was', 'and', 'is', 'this']
3	['it', 'i', 'the', 'not', 'a', 'to', 'and', 'my', 'is', 'for']
4	['i', 'it', 'the', 'not', 'this', 'and', 'to', 'in the mail', 'the mail', 'mail']
5	['it', 'i', 'the', 'a', 'and', 'to', 'not', 't waste', "don' t waste", 'this']
6	['it', 'i', 'the', 'to', 'and', 'is', 'not', 'a', 'this', 'for']
7	['it', 'thought it', 'thought', 'i thought', 'i thought it', 'i', 'to', 'the', 'a', 'not']
8	['it', 'the', 'i', 'a', 'to', 'not', 'is', 'this', 'and', 't']
9	['it', 'the', 'to', 'i', 'not', 'a', 'and', 'is', 'of', 'this']
10	['it', 'the', 'i', 'to', 'and', 'not', 'is', 'a', 'this', 'but']
11	['it', 'i', 'the', 'to', 'a', 'not', 'and', 'is', 'for', 'this']
12	['it', 'i', 'the', 'to', 'as', 'and', 'is', 'not', 'a', 'this']
13	['it', 'i', 'the', 'not', 'and', 'was', 'a', 'to', 'my', 'is']
14	['it', 'i', 'the', 'a', 'to', 'and', 'not', 'is', 'of', 'this']

As confirmed by the coherence score, the success of the LDA algorithm is very low when limited to 30 topics. It could not give almost any desired result. When the data in Table 3 is analysed in detail, it is observed that it is always stuck in similar classifications and gives bad results. However, as seen from the coherence score, there is a significant increase in the consistency score when the LDA algorithm is examined with more than 30 subjects.

**Table 4.** First 15 topic of 30 topic NMF Results.

Topic	Topic keywords
0	['use', 'to use', 'to', 'use it', 'to use it', 'easy', 'easy to', 'it', 'easy to use', 'i use']
1	['not work', 'work', 'did not work', 'did', 'did not', 'it did', 'not', 'it did not', 'work for']
2	['waste of', 'waste', 'of money', 'waste of money', 'money', 'a waste', 'a waste of', 'of']
3	["doesn", "doesn' t", 't', "it doesn", "it doesn' t", "doesn' t work", 't work', 'work', 'it']
4	['was', 'it was', 'it', 'was a', 'was not', 'it was a', 'i was', 'it was not', 'a', 'and it was']
5	['you', 'what you', 'you pay', 'pay for', 'pay', 'you get', 'get what', 'you pay for']
6	['the', 'in', 'case', 'in the', 'the case', 'on', 'on the', 'for the', 'and the', 'is']
7	['do', 'do not', 'buy', 'not buy', 'do not buy', 'not', 'buy this', 'not buy this', 'this', 'i do']
8	['all', 'at', 'at all', 'work at', 'work at all', 'all i', 'at all i', 'not work at', 'it at', 'it at all']
9	['of', 'of the', 'out', 'out of', 'out of the', 'the', 'box', 'a', 'the box', 'of the box']
10	['phone', 'my phone', 'on', 'on my', 'in my', 'phone and', 'the phone', 'and', 'charge']
11	['did', 'did not', 'fit', 'not', 'not fit', 'i did', 'i did not', 'did not fit', 'not like', 'did not like']
12	['back', 'get', 'my money', 'money back', 'my money back', 'money', 'my', 'i', 'get my']
13	['for', 'for my', 'bought', 'i bought', 'a', 'for a', 'and', 'it for', 'my', 'this']
14	['does not', 'it does', 'it does not', 'does not work', 'not work', 'work', 'does not fit']

When the NMF topic outputs in Table 4 are analyzed, it is seen that many topics have meaningful keywords and directly reflect the topic content. In addition to the high coherence score compared to LDA, when people are evaluated, it can be seen that the grievance topics that are the target of the study have emerged. For example, the topics numbered 2 3 4 12 have keywords that directly point to the problem.

Table 2 shows BERTopic usually gave the reasonable result regardless of the number of topics. Table 5 shows that most of the results are significant when sorted with 15 topics. When the coherence results in the Table 2 are analyzed, it is seen that the results get better as results go up to a certain number of topics and slightly decrease as move away from the optimum value.

Table 3, Table 4, and Table 5 are the results obtained for LDA, NMF, and BERTopic algorithms with the number of topics 30 parameter. In the study, the number of meaningful topics in the results obtained with the parameter number 175 was also examined. The criterion for a meaningful topic is that the topic content directly refers to the user complaint. According to the results obtained in this context, 16 topics out of 125 BERTopic were shaped directly related to customer complaints. In the LDA algorithm, 11 topics out of 175 were shaped directly related to customer complaints. Finally, 18 out of 175 topics in the NMF algorithm were shaped directly on customer complaints. Results includes average numbers. Since the aim of the study was to detect user complaints, the topics shaped around the product were not considered as success criteria in the study.



**Table 5.** BERTopic Results in 30 Titles

Topic	Topic keywords
0	['book', 'this book', 'read', 'the', 'this', 'was', 'it', 'the book', 'of', 'to']
1	['work', 'it', 'not work', 'not', 'did not work', 'did', 'did not', 'to', 'this', 'work at']
2	['case', 'the case', 'the', 'this case', 'it', 'is', 'case is', 'and', 'this', 'not']
3	['sound', 'ear', 'the', 'and', 'hear', 'is', 'the sound', 'the ear', 'it', 'they']
4	['cable', 'tv', 'hdmi', 'to', 'it', 'work', 'this', 'my', 'this cable', 'the']
5	['coffee', 'cup', 'tea', 'this', 'is', 'the', 'and', 'cups', 'taste', 'to']
6	['great', 'it', 'product', 'and', 'is', 'for', 'price', 'this', 'good', 'use']
7	['they', 'them', 'these', 'are', 'they are', 'and', 'they work', 'work', 'of', 'not']
8	['smell', 'the smell', 'this', 'it', 'the', 'smells', 'like', 'perfume', 'not', 'is']
9	['hair', 'my hair', 'my', 'it', 'and', 'this', 'hair and', 'not', 'hair it', 'to']
10	['phone', 'this phone', 'phone is', 'the phone', 'it', 'is', 'the', 'to', 'this', 'and']
11	['yet', 'it', 'here', 'order', 'and', 'was', 'to', 'long', 'not', 'still']
12	['it', 'love', 'she', 'love it', 'for', 'and', 'my', 'loves', 'loves it', 'for my']
13	['kindle', 'my kindle', 'fire', 'kindle fire', 'the kindle', 'my', 'it', 'the', 'fit', 'not']
14	['broke', 'it broke', 'it', 'after', 'broken', 'the', 'of', 'for', 'and', 'week']

## 5 Conclusion

Although it generally gives higher results in the coherence score, it cannot be said that the BERTopic model will always be better than LDA and NMF. This may vary depending on the data processed, time and many other criteria. LDA is one of the most widely used modelling methods. The reason for this is that it requires less resources than BERTopic, so it may not make sense to use BERTopic if there is not much time and powerful hardware. As a result of the study, it was seen that BERTopic, with the effect of using transformers embeddings, can separate topics well and shows a strong performance in topic modeling. In this study, the Amazon dataset was used and the comments included very short comments that contained noise or no meaning. However, the fact that the algorithm works well despite these negative factors is an indication that this study can be applied in many different fields. As an example, these results from topic modeling can be reflected back to the contents and turned into a classification problem, and new comments in the future can be categorized directly.

**Acknowledgment.** This study is supported by Artiwise Software Technologies.





## References

1. Fang, X., Zhan, J.: Sentiment analysis using product review data. *J. Big Data* **2**(1) (2015)
2. Gutiérrez, G.V.A.: A comparative study of NLP and machine learning techniques for sentiment analysis and topic modeling on amazon. *Int. J. Comput. Sci. Eng.* **9**(2), 159–170 (2020)

3. Egger, R., Yu, J.: A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify twitter posts. *Front. Sociol.* (2022)
4. Koruyan, K.: BERTopic Konu Modelleme Tekniği Kullanılarak Müşteri Şikayetlerinin Sınıflandırılması. *İzmir Sosyal Bilimler Dergisi* **4**(2), 66–79
5. Grootendorst, M.: BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv preprint [arXiv:2203.05794](https://arxiv.org/abs/2203.05794) (2022)
6. Sangaraju, V.R., Bolla, B.K., Nayak, D.K., Kh, J.: Topic Modelling on Consumer Financial Protection Bureau Data: An Approach Using BERT Based Embeddings. arXiv preprint [arXiv:2205.07259](https://arxiv.org/abs/2205.07259) (2022)
7. Abuzayed, A., Al-khalifa, H.: BERT for Arabic topic modeling: an experimental study on BERTopic technique. *Procedia Comput. Sci.* **189**, 191–194 (2021)
8. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003)
9. Févotte, C., Idier, J.: Algorithms for nonnegative matrix factorization with the  $\beta$  - divergence. *Neural Comput.* **23**(9), 2421–2456 (2011.)
10. Hoyer, P.: Non-negative matrix factorization with sparseness constraints. *J. Mach. Learn. Res.* **5**(9) (2004)
11. Donoho, D., Stodden, V.: When does non-negative matrix factorization give a correct decomposition into parts? *Adv. Neural Inf. Process. Syst.* **16** (2003)
12. McKinney, W.: Pandas: a foundational Python library for data analysis and statistics. *Python High Perform. Sci. Comput.* **14**(9), 1–9 (2011)
13. Lin, X., Boutros, P.C.: Optimization and expansion of non-negative matrix factorization. *BMC Bioinform.* **21**(1), 1–10 (2020)
14. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. *Nature* **401**(6755), 788–791 (1999)
15. Keita, Z.: Towards data science, Meet BERTopic BERT’s Cousin For Advanced Topic Modeling. Accessed 21 Nov 2022
16. Zvornicanin, Enes. When Coherence Score is Good or Bad in Topic Modeling. <https://www.baeldung.com/cs/topic-modeling-coherence-score>. Accessed 21 Nov 2022
17. Reimers, N., Gurevych, I.: Sentence-bert: sentence embeddings using siamese bert-networks. arXiv preprint [arXiv:1908.10084](https://arxiv.org/abs/1908.10084) (2019)
18. Goyal, C.: Step by Step Guide to Master NLP – Topic Modelling using NMF. <https://www.analyticsvidhya.com/blog/2021/06/part-15-step-by-step-guide-to-master-nlp-topic-modelling-using-nmf>. Accessed 21 Nov 2022



# Optimal Parameter Selection of Latent Dirichlet Allocation to Determine the Emerging Topics in Hydrology Domain

Sila Ovgu Korkut<sup>1,2</sup> , Aytug Onan<sup>3</sup> , Erman Ulker<sup>4</sup> ,  
and Femin Yalcin<sup>1</sup> 

<sup>1</sup> Department of Engineering Sciences, Izmir Katip Celebi University,  
35620 Izmir, Turkey

{[silaovgu.korkut](mailto:silaovgu.korkut@ikcu.edu.tr), [femin.yalcin](mailto:femin.yalcin@ikcu.edu.tr)}@ikcu.edu.tr

<sup>2</sup> Department of Translational Oncology, Dokuz Eylul University,  
35330 Izmir, Turkey

<sup>3</sup> Department of Computer Engineering, Izmir Katip Celebi University,  
35620 Izmir, Turkey

[aytug.onan@ikcu.edu.tr](mailto:aytug.onan@ikcu.edu.tr)

<sup>4</sup> Department of Civil Engineering, Izmir Katip Celebi University,  
35620 Izmir, Turkey

[erman.ulker@ikcu.edu.tr](mailto:erman.ulker@ikcu.edu.tr)

**Abstract.** In the new digital age, the determination of emerging topics has become a central issue for academia. Latent Dirichlet Allocation (LDA) method, a key mechanism for determining trends, has long been a method of great interest in a wide range of fields. However, the probabilistic structure leads to a serious effect on the score not only by changing the parameters but also from trial to trial for the fixed parameters. This study describes the implementation of the LDA method for exploring trend topics in the hydrology domain. Several parameters like the portion of the corpus, the number of topics as well as hyperparameters of the LDA method,  $\alpha$  and  $\beta$ , have been considered. The emerging topics of the field have been obtained using the parameters which attained the highest mean coherence score.

**Keywords:** Latent Dirichlet Allocation · Text Analytics · Hyperparameters · Topics in Hydrology

## 1 Introduction

The rapid advancements in ubiquitous technology have brought about significant and inevitable changes to academia [1]. One of the most noticeable impacts is the exponential increase in the number of academic journals, which has led to a dramatic increase in the volume of publications. This has created challenges

---

Supported by Center of Scientific Research Projects of the Izmir Katip Celebi University [Grant Number: 2022-GAP-MUMF-0029].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024  
F. P. García Márquez et al. (Eds.): ICCIDA 2023, SCI 1145, pp. 32–42, 2024.  
[https://doi.org/10.1007/978-3-031-53717-2\\_4](https://doi.org/10.1007/978-3-031-53717-2_4)

for researchers to keep up with the latest research in their field. Another issue that has emerged is the untraceability of research articles in the digital era, making it challenging to trace the evolution of research over time. Additionally, the unpredictable shifts in research topics and interests have made it difficult for researchers to stay current with emerging trends. As a result, addressing any of these issues has significant implications for academia.

Emerging topic identification is a process of discovering and characterizing new or previously under-explored themes in a particular domain, such as scientific research, social media, or news articles. It involves using data analysis techniques, such as topic modeling or text mining, to identify patterns, trends, and relationships within large collections of text data. The goal of emerging topic identification is to help researchers or practitioners stay up-to-date with the latest developments in their field and to identify new research directions or opportunities [2]. In this paper, the determination of emerging topics playing a vital role in managing the time of researchers has been addressed.

The Latent Dirichlet Allocation (LDA) model is a widely used probabilistic method for identifying topics in a text corpus [3]. Numerous studies have investigated the effectiveness of the LDA model for capturing different facets of topics in academia. The literature on LDA implementation in academic research can be summarized as follows: Griffiths and Steyvers used the LDA model to analyze abstracts of papers gathered from PNAS from 1991 to 2001 to determine the hot topics in [8], Linstead *et al.* have analyzed the software evolution via the LDA model in [12] by carrying out multiple versions of the Eclipse and ArgoUML projects. Kim and Chen [15] have investigated the research topic trends in the field of Engineering Management. IEEE Transactions on Engineering Management journal has been the target for gathering the articles excluding reviews, editorial notes, and guidelines in 20 years time span from 1998 to 2017. The 922 collected articles from the IEEE Xplore database have been analyzed and the authors have presented 40 topics including hot topics, steady topics, and cold topics taking into account the alteration of the number of articles over the years. More recently, in the study of Han [11] a diversity of research topics and trends in library and information science (LIS) has been exhibited via the LDA model. The articles were gathered from the Scopus database. The time interval, from 1996 to 2019, has been divided into five intervals covering at most the span of 4 years. Moreover, the authors have performed the LDA model to conduct a literature analysis on COVID-19, MERS, and SARS coronavirus diseases, [13]. To do so, 7909 full-text documents in English were accumulated from PubMed Central, bioRxiv, medRxiv, and others. The results of that study have been to return the most relevant terms appearing in coronavirus-related research, present a guide for the medical community for literature, and visualize a new conceptual network to provide the similarities and differences between topics. Furthermore, the debate between full-text or abstract analyses is discussed in [9] taking into account an expert opinion.

Due to the stochastic nature of the LDA model, it can produce different topics for the same corpus even if the parameters are held constant. The result-

ing variability can be challenging to manage, especially when comparing different models. One can address this issue by running multiple trials of the LDA model and averaging the results to provide a more robust representation of the underlying topics and reduce the impact of the stochasticity of the model. In addition to the variability of the LDA model, the choice of hyperparameters can also have a significant impact on the results. The study in [10] has explored the effect of different Dirichlet priors on topic coherence, both symmetrical and asymmetrical, in scientific research articles related to the domain of fisheries. The authors in [10] have found that the choice of hyperparameters can influence the number, coherence, and interpretability of the resulting topics. Therefore, it is crucial to carefully choose the hyperparameters to obtain meaningful and accurate results from the LDA model. Furthermore, recently published articles, [4, 14], have explored the application of LDA in a variety of fields beyond fisheries, including medicine, finance, social sciences, and mathematical expressions. These studies have demonstrated the versatility and effectiveness of LDA in capturing the underlying topics in diverse domains. The use of LDA in these fields has helped to uncover emerging research areas, identify important trends, and provide valuable insights into complex systems.

This paper assesses to find out the optimal parameter selection of the LDA model on a keyword in academia. To that aim, the “hydrology” keyword, which is already associated with the trending climate change, has been chosen. It is worth noting that an experimental study has been discussed in the current study. To do so, various parameters of the LDA model such as the number of topics, portion of the corpus, and controlling parameters of document-topic distribution and topic-word distributions have been taken into account. To enrich the scope of the layout all parameters are discussed for various combinations of words.

The remaining part of this study is composed of three main sections as follows: Sect. 2 begins with laying out the methodology of the research by giving the details of the data collection, inclusion and exclusion criteria of the data pre-processing, and the theoretical information about the Latent Dirichlet Allocation. It is followed by Sect. 3 to present what stands out from the results. In addition to the depicted figures and tables, a detailed discussion has been provided. Furthermore, the attained trend topics concerning optimal hyperparameters have been discussed by a field expert. Section 4 summarizes the findings and all conclusion remarks of the research.

## 2 Methodology

### 2.1 Data Collection

Following the essential aim of the study, data is acquired from the bibliographic database of *Web of Science*<sup>1</sup> which provides bibliographic features of research

---

<sup>1</sup> Web of Science (Web of Knowledge) owned by Clarivate is a scientific platform providing articles indexed by itself (Science Citation Indexed-Expanded (SCI-E); Emerging Science Citation Indexed (ESCI), Social Science Citation Index (SSCI),

papers. The dataset was gathered by employing the “Hydrology” keyword on the time interval from January 2008 to May 2022. The shape of the initial dataset was (56951, 70) consisting of 56951 articles, conference proceedings, books, and editorial notes with 70 bibliographic information.

## 2.2 Inclusion and Exclusions of the Study

The scope of this study has been confined to “Article” and “Proceeding Paper” type items while excluding other research items such as review articles, editorial notes, biographical items, and data papers. After basic preprocessing, the dataset was reduced to (53075, 20). To align with the scope of the study, only textual-related features such as “Article Title” and “Keywords” were filtered for analysis. These features are critical for attracting readers and providing a summary of the research, making them suitable for the study’s purposes. As a result, the final dataset used for analysis had a shape of (53075, 2). The textual data underwent further preprocessing, which included tokenization, stemming, removal of noise and stop words, and filtering tags. The authors used the Natural Language Toolkit (NLTK) [5] for these tasks and filtered verbs and conjugations. Additionally, stop words were extended by considering several words, including the stem of the main keyword, based on the guidance of field experts. Finally, the authors generated bi-gram and tri-gram phrase models using the *Gensim* library, see [6].

## 2.3 Latent Dirichlet Allocation

The Latent Dirichlet Allocation (LDA) is one of the prominent topic modeling methods which was first put forth in 2003 by Blei et al. [3]. The feature that makes this theory stand out from other probabilistic models is the interpretation of the exchangeability as conditionally independent and identically distributed. The main scope of the LDA model is to determine the significant members of the corpus and similar documents by assigning higher probabilities.

The process of the LDA model can be summarized as follows: Define a vocabulary with size  $|V|$  composed of the basic units of textual data (words) where the words are represented by 0 and 1 such that the  $k$ th word ( $w$ ) in the vocabulary  $w^k = 1$  and  $w^l = 0$  for  $l \neq k$ . Let  $\mathbf{w} = (w_1, w_2, w_3, \dots, w_N)$  stand for the document consisting of  $N$  words in an  $M$  document collection of the corpus such that  $D = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_M)$  and  $K$  denote the number of topics. For each word in a document, the process is implemented as follows: choose  $z_i \sim \text{Multinomial}(\theta)$  and  $w_i$  from  $p(w_i|z_i, \beta)$ , where  $\theta \sim \text{Dir}(\alpha)$  stands for the  $K$  dimensional random variable where  $\alpha$  represents document-topic density and the  $K \times V$  matrix  $\beta$  denotes topic-word density such that  $\beta_{ij} = p(w^j = 1|z^i = 1)$ . For a given  $\alpha$  and  $\beta$ , the joint distribution of a topic mixture  $\theta$ , a set of  $N$  topics  $\mathbf{z}$ , a set of  $N$  words  $\mathbf{w}$ , the probability of topics in the LDA model, [3], is identified by

---

Arts & Humanities Citation Index (AHCI)) with many features belonging to the article including citations, references, article title, abstract, etc. for various fields.

$p(\theta, \mathbf{z}, \mathbf{w}|\alpha, \beta) = p(\theta|\alpha) \prod_{i=1}^N p(z_i|\alpha) p(w_i|z_i, \beta)$ . For further details, we refer the interested reader to [3].

The LDA model represents documents as a random mixture of topics, with each topic described by a Multinomial distribution, making it a probabilistic model. To evaluate the performance of the LDA model for the specified textual data, we used the  $C_V$  coherence score. This score is a coherence measure that combines indirect cosine similarity with normalized pointwise mutual information, as described in [7]. However, being a probabilistic model, the LDA can return different topics and scores for the same corpus, which can lead to variability in the results. To obtain more reliable scores, we performed the LDA model 10 times for different values of the combination of  $(K, \alpha, \beta)$  and recorded the average coherence scores. To illustrate the methodology of this study, we present a graphical summary in Fig. 1. After generating the main corpus and dictionary for both bi-gram and tri-gram phrases, we performed the LDA model on the data and recorded the average coherence scores. This approach allowed us to assess the performance of the LDA model in identifying emerging topics in hydrology and to determine the optimal values of the hyperparameters  $(K, \alpha, \beta)$  for the model. This section ends up with Fig. 1 to illustrate the methodology of this study, schematically.

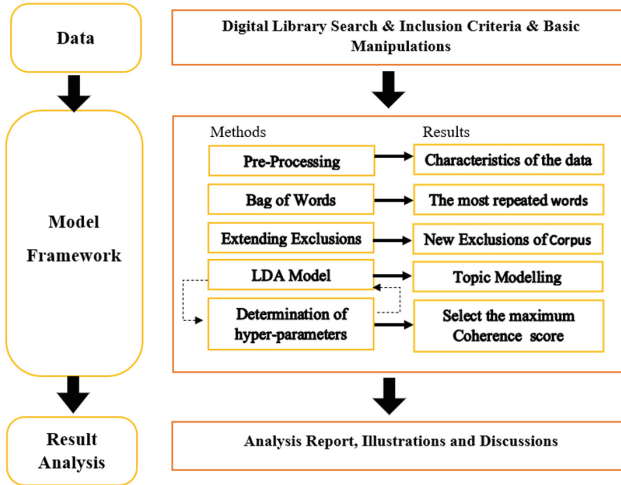


Fig. 1. The graphical summary of the methodology.

### 3 Results and Discussions

This section aims to pinpoint optimal values of the parameters of the LDA model taking the coherence scores ( $C_V$ ) into account. A grid search has been carried out for the values of  $\alpha$  and  $\beta$  as well as the number of topics. To do so, in addition to article titles keywords of 53075 research items related to the “Hydrology” keyword are fed to the LDA model. Along with “symmetric” and

“asymmetric” distribution choices, the values of  $\alpha$  have varied between [0.01, 1] by 0.3 increments. Likewise, the  $\beta$  parameter is considered on the interval [0.01, 1] with 0.3 step size as well as a “symmetric” distribution alternative. Moreover, the number of topics has varied from 3 to 7. Furthermore, various portions (50%, 75%, 100%) of the corpus have also been taken into account to see the effects of the portion of the corpus on the score of the LDA model. It is crucial to remind that the LDA model has been implemented 10 times to record the coherence score for each specified combination. Thereafter, the mean of those 10 records has been stored as the score of that parameter. The average of mean coherence scores with respect to each parameter has been given in the following tables.

**Table 1.** The average of the mean coherence scores of the LDA model for 10 trials with respect to the hyperparameter  $\beta$ .

$\beta$	Bi-gram	Tri-gram
0.01	0.4144	0.3923
0.31	0.4093	0.3884
0.61	0.4147	0.4012
0.91	0.4109	0.4064
symmetric	0.4085	0.3879

**Table 3.** The average of the mean coherence scores of the LDA model for 10 trials with respect to the portion of the corpus.

PC	Bi-gram	Tri-gram
100% Corpus	0.4133	0.4042
75% Corpus	0.4128	0.3950
50% Corpus	0.4085	0.3865

PC: portion of the corpus

**Table 2.** The average of the mean coherence scores of the LDA model for 10 trials with respect to the hyperparameter  $\alpha$ .

$\alpha$	Bi-gram	Tri-gram
0.01	0.4070	0.3887
0.31	0.4157	0.4009
0.61	0.4169	0.3975
0.91	0.4162	0.4002
asymmetric	0.3998	0.3908
symmetric	0.4136	0.3932

**Table 4.** The average of the mean coherence scores of the LDA model for 10 trials with respect to the number of topics.

NT	Bi-gram	Tri-gram
3	0.3542	0.3513
4	0.4011	0.3657
5	0.4061	0.4155
6	0.4458	0.4220
7	0.4506	0.4215

NT: number of topics

Tables 1, 2, 3 and 4 show the variability on  $\alpha$ ,  $\beta$  and the portion of the corpus have fewer impacts on the average of mean coherence scores than the number of topics. Moreover, it is apparent from these tables that the averages of mean coherence scores are increasing as both the number of topics and the portion of the corpus are increasing. However, there is no such obvious relation between the average of the mean coherence scores and the values of  $\alpha$  and  $\beta$ . It is crucial to say that the LDA model is performed by using the highest coherence score to obtain the topics in the field. Therefore, 5 combinations of the specified param-



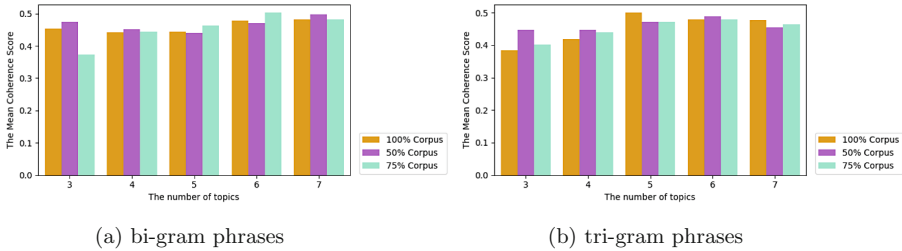
eters attained the highest mean coherence scores for both bi-gram and tri-gram phrases are listed in Table 5.

**Table 5.** The highest 5 mean coherence scores of the LDA model with 10 trials for various parameters.

No.	Bi-gram						Tri-gram					
	PC	NT	$\alpha$	$\beta$	MC	STD	PC	NT	$\alpha$	$\beta$	MC	STD
1	75%	6	symmetric	0.91	0.502660	0.004058	100%	5	asymmetric	0.91	0.499852	0.010802
2	50%	7	symmetric	0.61	0.498058	0.007774	50%	6	asymmetric	0.31	0.488762	0.004858
3	50%	7	0.01	0.61	0.493917	0.008434	100%	6	0.61	0.91	0.480124	0.003187
4	75%	6	0.01	0.91	0.487246	0.029724	75%	6	0.61	0.91	0.479961	0.000979
5	100%	7	0.31	0.61	0.482370	0.005494	100%	7	0.91	0.91	0.477731	0.002073

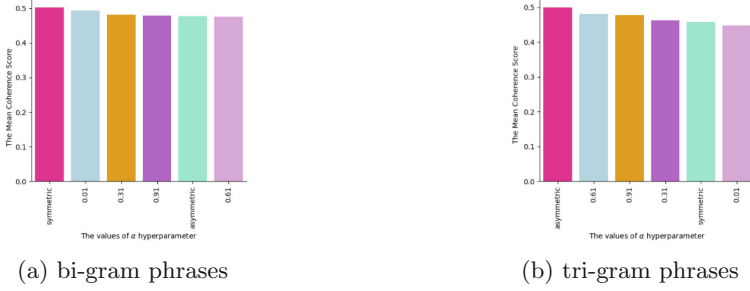
PC: Portion of the corpus, NT: Number of topics, MC: Mean of the measured coherence scores, STD: Standard deviation of the measured coherence scores

Contrary to expectations of Table 3–4, the highest mean coherence score, for the bi-gram case, has been recorded as 0.50 with a standard deviation of 0.0041 on 75% of the corpus with 6 topics where  $\beta = 0.91$  and  $\alpha$  is symmetric. On the other hand, for the tri-gram case, the recorded highest coherence score is almost 0.50 with a standard deviation of 0.0108 for the whole corpus with 5 topics where  $\beta = 0.91$  and  $\alpha$  is asymmetric. This implies the combination of the number of topics and the portion of the corpus can play a significant role in the mean coherence scores. To see the impact of that combination Fig. 2 has been presented. To get the depicted figure, the recorded list from 10 trials for each combination of parameters is grouped by the features of “The Number of Topics” and “The Portion of the Corpus” aggregating by the maximum value of the coherence scores.

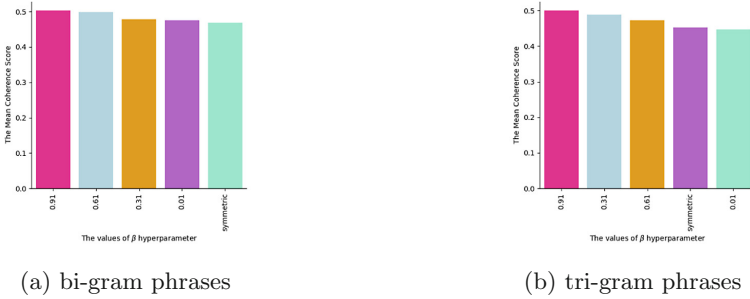


**Fig. 2.** Effects of the number of topics on the maximum value of the mean coherence scores of the 10-trial LDA implementation by clustering the portion of the corpus.

The highest mean score is recorded as 6 topics for the bi-gram whereas that value is achieved as 5 for the tri-gram case. Therefore, the results show that the values of the topics to attain the maximum coherence score can vary with respect to the combination of  $n$  sequence of words. Moreover, the functionality of



**Fig. 3.** Maximum value of the mean coherence scores versus various values of  $\alpha$  parameters.

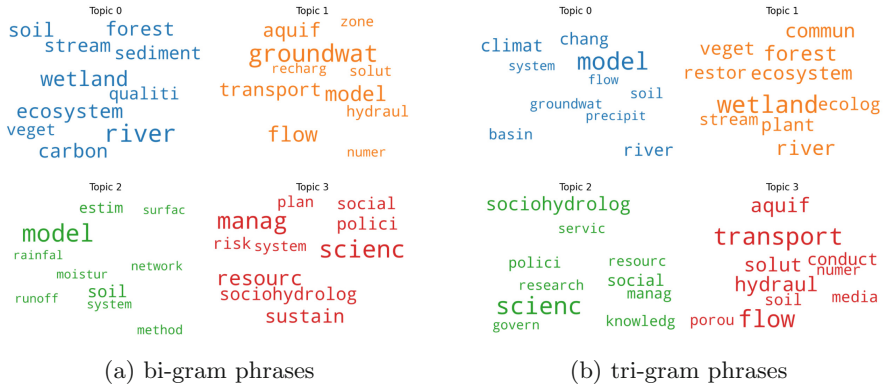


**Fig. 4.** Maximum value of the mean coherence scores versus various values of  $\beta$  parameters.

$\alpha$  and  $\beta$  parameters on the maximum value of the mean coherence score has been checked by grouping the recorded list with respect to the feature of “Number of Topics” and aggregated by the maximum value of the coherence scores. The results are demonstrated in Fig. 3 and Fig. 4, respectively.

It can be deduced from Figs. 3 and 4 that the hyperparameters  $\alpha$  and  $\beta$  are not as effective as the other parameters on the mean coherence scores. However, the apparent variability in tri-gram associations is relatively greater than those in bi-gram. Before ending the section, the suggested topics of the specified domain for the hyperparameters whose values achieved the highest coherence scores for bi-gram and tri-gram choices are illustrated in Fig. 5a and Fig. 5b, respectively.

The prominent topics in Fig. 5a can be interpreted as follows: Topic 0 describes the premise of publications related to the effects of surface water on the quality of the ecosystem via analyzing the flow region and sediment transport concerning vegetation. Additionally, the studies based on the numerical models of hydraulic transport in groundwater in aquifers are highlighted in Topic 1. In Topic 2 the estimation method dealing with the entire hydrological cycle network formed on the basin comes to the fore. Furthermore, it can be inferred from Topic 3 that the studies in water policy and management, which should be con-



**Fig. 5.** Topics extracted from the LDA model selecting the hyperparameters which return the highest coherence score

sidered socially for sustainable water resources become popular in recent years. On the other hand, what emerges from the results of tri-gram analyses which are presented in Fig. 5b is as follows: Topic 0 depicts the effects of climate change on the hydrological cycle, such as the impact of climate change on precipitation, surface runoff, and groundwater flows. Moreover, Topic 1 underlines the studies on restoring ecosystems of communities living in river basins through conscious forestation and vegetation. While the importance of informing the public of the state’s water resources policies and management systems can be deduced from Topic 2, Topic 3 is the solution of the numerical modeling of the underground water flow in the aquifer together with the hydraulic conductivity and porosity of the soil.

## 4 Conclusion

This study was designed with the aim of assessing the effects of the parameters on the LDA model scores in the “Hydrology” domain. The various combinations of the values of the main parameters such as the number of topics, the portion of the corpus, and the values of  $\alpha$  and  $\beta$  have been studied on both bi-gram and tri-gram phrases via the **Gensim** library. Taking the probabilistic behavior of the LDA model into account, to obtain more reliable scores each combination has been implemented 10 times, and the mean coherence scores and their standard errors have been recorded. The results show that various combinations of the parameters play an important role in the mean coherence scores. Furthermore, the possible trend topics from bi-gram and tri-gram phrases have been presented for the parameters for which the highest coherence scores have been attained. Taking the topics attained from both phrases together, these results provide important insights into the main trends in the field: Handling the hydrological systems from a mathematical point of view has still attractive

among researchers. Moreover, the studies on forestry and vegetation due to climate change are popular, and Social awareness and governmental policies about water resources protection have an increasing impact on the field. After consulting the experts' decisions, it is concluded that the LDA method has succeeded to return to the trend topics of Hydrology. However, it is worth noting that due to the stochasticity of the method, the coherence score of the LDA method can vary even for fixed parameters.

## References

1. Bornmann, L., Mutz, R.: Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references. *J. Assoc. Inf. Sci. Technol.* **66**, 2215–2222 (2015). <https://doi.org/10.1002/asi.23329>
2. Small, H., Boyack, K.W., Klavans, R.: Identifying emerging topics in science and technology. *Res. Policy* **43**(8), 1450–1467 (2014). <https://doi.org/10.1016/j.respol.2014.02.005>
3. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003)
4. Chauhan, U., Shah, A.: Topic modeling using latent Dirichlet allocation: a survey. *ACM Comput. Surv. (CSUR)* **54**(7), 1–35 (2021). <https://doi.org/10.1145/3462478>
5. Hardeniya, N., Perkins, J., Chopra, D., Joshi, N., Mathur, I.: *Natural Language Processing: Python and NLTK*. Packt Publishing, Birmingham (2016)
6. Rehurek, R., Sojka, P.: *Gensim-python framework for vector space modelling*. NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic, vol. 3, no. 2 (2011)
7. Röder, M., Both, A., Hinneburg, A.: Exploring the space of topic coherence measures. In: *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, pp. 399–408 (2015). <https://doi.org/10.1145/2684822.2685324>
8. Griffiths, T.L., Steyvers M.: Finding scientific topics. *Proc. Natl. Acad. Sci.* **101**, 5228–5235 (2004). <https://doi.org/10.1073/pnas.0307752101>
9. Syed, S., Spruit, M.: Full-text or abstract? Examining topic coherence scores using latent dirichlet allocation. In: *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 165–174 (2017). <https://doi.org/10.1109/DSAA.2017.61>
10. Syed, S., Spruit, M.: Selecting priors for latent dirichlet allocation. In: *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*, pp. 194–202. Laguna Hills, CA, USA (2018). <https://doi.org/10.1109/ICSC.2018.00035>
11. Han, X.: Evolution of research topics in LIS between 1996 and 2019: an analysis based on latent dirichlet allocation topic model. *Scientometrics* **125**, 2561–2595 (2020). <https://doi.org/10.1007/s11192-020-03721-0>
12. Linstead, E., Lopes C., Baldi, P.: An application of latent dirichlet allocation to analyzing software evolution. In: *2008 Seventh International Conference on Machine Learning and Applications*, pp. 813–818. San Diego, CA, USA (2008). <https://doi.org/10.1109/ICMLA.2008.47>
13. Cheng, X., Cao, Q., Liao, S.S.: An overview of literature on COVID-19, MERS, and SARS: using text mining and latent dirichlet allocation. *J. Inf. Sci.* **48**(3), 304–320 (2022). <https://doi.org/10.1177/0165551520954674>

14. Sakshi, V.K.: Recent trends in mathematical expressions recognition: an LDA-based analysis. *Expert Syst. Appl.* **213**(Part B), 119028 (2023). <https://doi.org/10.1016/j.eswa.2022.119028>
15. Kim, J.H., Chen, W.: Research topic analysis in engineering management using a latent dirichlet allocation model. *J. Ind. Integr. Manag.* **3**(4), 1850016 (2018). <https://doi.org/10.1142/S2424862218500161>



# Battery Management System-Based Fuzzy Logic

K. S. Jithin Mohan<sup>1</sup>(✉) and S. Paul Sathiyam<sup>2</sup>

<sup>1</sup> Electrical and Electronics Engineering, Karunya Institute of Technology and Science,  
Coimbatore, Tamil Nadu, India

jithinmohan03@gmail.com

<sup>2</sup> EEE Department, Karunya Institute of Technology and Science, Coimbatore, Tamil Nadu,  
India

**Abstract.** To solve the issue of battery charge-discharge and associated damage brought on by incorrect estimates of the battery efficiency, fuzzy logic is used to define a new quantity known as the Energy storage system (ESS), which is based on the battery state, state of charge (SOC), and state of health (SoH). A battery management system (BMS) technique is necessary for energy storage systems (ESSs) for ageing increases a battery's internal resistance and reduces its capacity. To control the battery state using fuzzy logic, in this paper, a formula for calculating battery efficiency is proposed. The charging time, charging current, and battery capacity are all factors in the proposed fuzzy logic-based battery efficiency estimation formula. The findings show that the ESS is used by the fuzzy logic battery management system to determine battery efficiency. The battery efficiency is also decreased by using a defect diagnosis algorithm to construct a safe system when charging and discharging. Applying the proposed BMS algorithm in a 3-kW ESS shows that it is valid.

**Keywords:** Energy storage system · state of charge · fuzzy logic · state of health · battery management

## 1 Introduction

The current energy requirements of customers will be met by the electric-powered grid. Paint generators to satisfy loads with wide height-to-base variances, however, are particularly challenging due to the large need for power modifications during the day and year [1]. Power suppliers must always maintain a decent enough capacity to meet real-time demand and enough hooked-up power functionality to meet height requirements. Frequently, operating capabilities at 20% above expected demand and, at maximum, using 55% of the complete generation functionality for a year are required to meet one's needs. By using a power garage, technological assets can operate at full capacity as opposed to being scaled up or down while a power garage is paying for variations in the name. Applications that are received from power storage units on the electric grid have a wide range of requirements. A seasonal power storage plant that requires megawatt-hours of capacity stored for extended periods is required in a few remote locations [2, 3].

The most straightforward operation of every battery%, supervisory tool management (SSC), and a battery management tool (BMS) are required for BESSs to show and maintain them stable [4]. Batteries are dynamic and constantly perform the equilibrium nation outdoors throughout cycling. Furthermore, the scenario will become worse for intercalation-primarily based complete storage systems (such as Li chemistry), which perform as closed gadgets with few quantifiable countrywide variables, making it difficult to correctly screen the states of the battery and hold stable operation. Furthermore, even beneath flawless ordinary functioning, the battery packs of a BESS will ultimately lose their capacity. This degradation can be sped up by using immoderate charging patterns, expanded temperature (each ambient and operating), overcharging, or undercharging. Battery packs are managed via way of means of a simple BMS to be able to meet the call for electricity best [5, 6].

The UN has placed a strong emphasis on requiring renewable electricity to reduce carbon emissions. In line with this, numerous nations are utilizing renewable energy to generate electricity. The deployment of ESS had previously attracted a lot of interest [7]. Tracking those ESS has made it possible for battery-control structures to enforce irregularities and enable ESS fault detection, though. The global market for battery-control systems for various applications is expected to grow at a compound annual growth rate (CAGR) of 54.8% as a result of wireless separation into two main connection-based subsystems [8, 9]. With the need to eliminate wires and utilize the IoT, wireless battery-control structures are quickly gaining popularity. A battery-control device also ensures that a tool's internal structure will have pre-configured corrective actions on rare occurrences. A battery control device also confirms the proper method for regulating a device's temperature because the temperature affects the power-consumption profile [8, 10].

The car enterprise has superior significantly because of several technological trends which have introduced passenger and pedestrian protection nearer to reality. However, the growth withinside the number of automobiles on the street is to blame for the pointy upward push in pollutants degrees in city areas [11]. According to the European Union, the transport enterprise is responsible for around 27% of all carbon dioxide (CO<sub>2</sub>) emissions, while car transport is responsible for more than 70% of the emissions [12]. Due to their capacity to decrease environmental pollutants, preserve fossil fuels, and reduce carbon emissions and global warming challenges, electric-powered automobiles (EVs) have attracted significant hobby and popularity withinside the area to deal with such issues. In addition to pollutants degrees, EVs provide diesel-primarily based motors a promising opportunity in phrases of simplicity, dependability, comfort, and performance. However, the vast use of EVs necessitates the battery garage system's (BSS) proper features and diagnosis in phrases of molecular battery monitoring, charge-discharge control, molecular balancing, temperature control, and power regulation [13, 14].

## 2 Related Work

Globally, studies on the battery era for electric-powered automobile programs are developing remarkably to cope with the troubles of carbon emissions and global warming. Electrical motor performance relies upon the right evaluation of essential elements, in

addition to the functionality and diagnosis of the battery garage tool. However, terrible battery garage tool monitoring and protection tactics can bring about vital troubles consisting of battery overcharging, over-discharging, overheating, molecular unbalance, thermal runaway, and hearth dangers. A sturdy battery control tool performs an essential position in improving battery overall performance via way of means of appearing particular monitoring, charging-discharging manipulation, warm temperature manipulation, battery protection, and protection good ways to cope with the troubles. The reason for this study is to [1] offer an intensive assessment of unique foxy strategies and battery manipulation tool manipulation strategies utilized in electric-powered automobile systems. Accordingly, the assessment assesses the smart algorithms for battery nation estimate in phrases of their features, structure, configuration, accuracy, benefits, and drawbacks. The examination additionally examines the numerous controllers used for battery heating, cooling, equalization, and protection, highlighting numerous categories, traits, goals, successes, and drawbacks. The predominant demanding situations and difficult instances in phrases of computation complexity, execution troubles, and different inner and outside elements are recognized. Finally, destiny potentialities and recommendations are offered as a good way to lay out an environmentally aware set of recommendations and controllers for the development of a complicated battery manipulate device for destiny sustainable electric powered automobile initiatives.

Comparing the fitness region of the battery well is of critical relevance for electric vehicles to ensure riding protection and prevent capacity disasters. Through a rigorous assessment of the to-be-had literature on the state of health estimate strategies, this looks seek [2] to function as a valuable aid for researchers and practitioners. These techniques may be categorized into experimental and model-primarily based overall estimation techniques. To have a look at battery aging strategies and offer theoretical aid for model-primarily based tactics, experimental strategies are utilized in a lab setting. To understand the state of fitness estimation, model-based complete estimate procedures use a battery model to identify the factors that have specific associations with the battery getting older level. Tactics for determining the fitness country of the batteries are specified in further detail, and the accompanying merits and limitations of those strategies are assessed in this work based on an analysis of a sizable body of literature. Conclusions for those tactics and opportunities for the enhancement of fitness country estimation are given at the top of the study.

The evolution of BMS from a minor tracking contains several functions one is expanded with the aid of using improvements in the battery age. To apply the BMS's mighty management, it's far important to set up a battery version. It can be correct and powerful to mirror the behavior of the battery degree to the car with the usage of an extra whole and quicker battery version. Based on this, [3] some important technologies, consisting of battery state estimate, energy equalization, temperature management, and fault diagnosis, are superior as a way to make certain battery safety, power, and durability. In addition, verbal exchange among BMS and automobile controllers, motor controllers, and different gadgets is important for maximizing using and boosting car performance. As said earlier, a synergistic and collaborative BMS is the idea for intelligent, electric-powered, networked, and shared inexperienced power vehicles. The studies and development (R&D) of the Multiphysics version simulation and the Multifunction



Included BMS Era are consequently mentioned in this paper. Furthermore, state-of-the-art work generation and pertinent studies are devoted to advancing BMS coordination and synergy in addition to promoting the innovation and optimization of the maximum latest electric-powered car era.

The extraordinary environmental repercussions of gasoline/diesel-primarily based total vehicles over the latest few long time are the purpose for the developing recognition of electric vehicles (EVs). EVs were commercialized in some locations around the world, however, they may be now completely commercialized. Even eleven though utilizing EVs has many benefits, there are numerous difficult circumstances, together with variety anxiety, gradual charging, or the effectiveness/charge of the battery. The discipline of battery management structures has gone through an in-depth overview from the years 2006 to 2020. (BMS). The numerous roles, [15] benefits, and downsides of the strategies utilized in BMS for molecular balancing, thermal management of the battery, the safety of the battery towards overvoltage and overcurrent, estimation of State of Health (SOH), and estimation of the country of charge (SOC), are mentioned on this paper. Additionally, the crucial gaps have been identified, and a framework for an inexperienced BMS layout changed suggested. Also protected in the works of the one is the software program for modern-day, wise technology, together with the Digital Dual of the Battery Pack, cyber-bodily structures, battery swapping technology, non-awful battery testing, self-reconfigurable batteries, and responsible automatic recycling and reuse. This statement offers important gaps, modern-day technology, and a framework that specialists and researchers may make use of to expand completed structures that encompass stronger BMS, in real time.

This paper [16] proposes a fuzzy, common region revel in-primarily based electricity control system (EMS) for microgrids with a mixed battery and hydrogen ESS, which ensures electricity stability under the burden call the same time as taking into consideration the development of the micro grid's universal overall performance from a technical and financial factor of view. As is known, the studies community is developing more and more interest in renewable electricity-primarily based microgrids as they may be essential to the venture of growing the imminent electricity transition model. For planning the operation of renewable electricity-primarily based microgrids, a reliable forecasting device and inexperienced dispatching manipulation among the to-be-had and favored electricity are necessary. Modern EMS is stimulated through the bushy now no longer unusual region reveling in the controller (FLC), which has precise blessings over different controllers. Knowing the plant's model is not necessary, and understanding the linguistic regulations that make up its inference engine is simple. These laws may encompass professional knowledge, which simplifies the generally tough micro-grid manipulation. The higher EMS has passed through a pressure take a look at that has validated its appropriate behavior. A residential-kind profile has been hired for it in a true microgrid. The advanced fuzzy now no longer unusual region revels in-primarily based EMS can meet each technical (to grow the devices' lifespan) and financial (looking for pleasant profitability and efficiency) said requirements, similar to responding to the required load call. A professional may want to offer this carrier, relying upon the micro grid's role and profile calls to complete.

### 3 Proposed Method

To produce the desired outcome, the fuzzy system must take into account all of the factors affecting the EV's energy consumption; the energy demand, SOC, and SoH are the system's inputs because they provide analysis of numerous factors at once and system

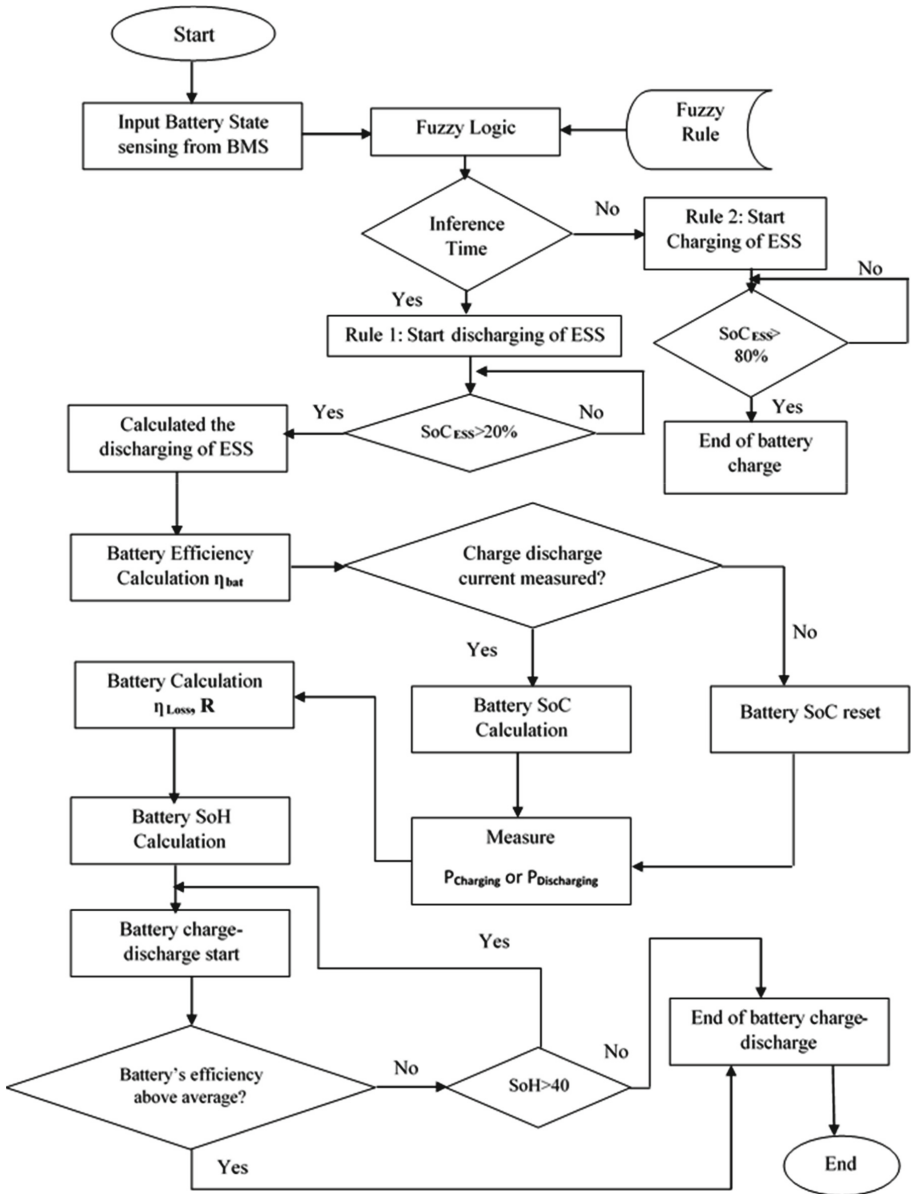


Fig. 1. Overall diagram for the proposed methodology

feedback. Because the ESS, which charges and discharges electricity from a battery, is directly related to the battery’s overall performance, it is crucial to have a BMS that manages, secures, and interacts with the outside world. Since one of the issues with non-linear batteries is that they have internal properties that change depending on how many charge-discharge cycles they have through, the main objective of a BMS is to correctly record these changes. In this research, a method for improving battery performance and safety is proposed. Fuzzy logic is used to construct a battery performance formula, which is then used to compute and predict the SoC and SoH of the battery based on the performance degradation that occurs during battery use. This study also suggests a fuzzy logic-based battery management system for stopping the use of the battery and identifying errors, which is mostly based on battery performance computation. To manage the battery kingdom using fuzzy logic, this research proposes a battery efficiency calculation formulation. The suggested fuzzy logic-based formulation for calculating battery efficiency makes use of the battery capacity, SoC, SoH, and charging and discharging times. Figure 1 illustrates the overall diagram of the proposed methodology.

In an ESS system, the suggested algorithm was put into practice. The ESS is made up of two parts: a battery system and a power conversion system (PCS). A battery and a BMS made up the battery system. The ac-dc converter on the PCS was a two-level, highly efficient, and manageable device, in contrast to the dc-dc converter, which was a full-bridge device (Fig. 2).

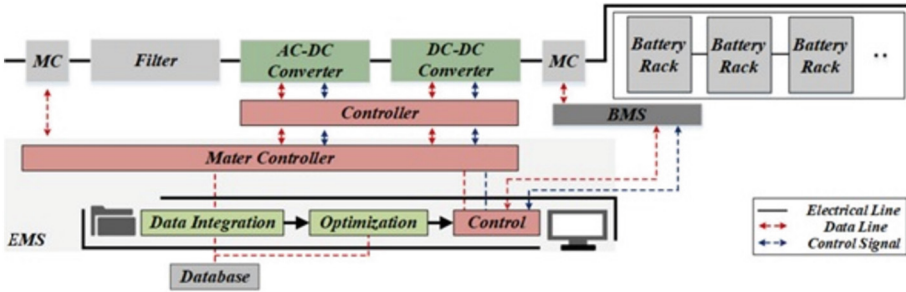


Fig. 2. Proposed ESS structure illustration

### 3.1 ESS Considering Battery Efficiency

The ESS helped the grid by discharging during the day and charging at night. While the charge-discharge method was being developed utilizing the BMS set of rules, the recommended methodology should sense the battery state using CAN verbal communication. If the SoC of the ESS battery dropped below 80%, charging started; if it rose above 80%, charging ceased. The operational SoC is defined in this document as that described between 20% and 80%; the person may specify the battery SoC’s least and maximum values. The ESS continued to discharge if the grid insisted on more electricity after charging. Furthermore, the ESS employed a set of guidelines that prohibited battery discharge when the SoC went below a threshold of 20%. The EMS set of guidelines for

a battery's charge-discharge technique ensures remarkably high safety when employed alongside the BMS [17].

### 3.2 Proposed Fuzzy Logic-Based BMS Algorithm

The main objective of a BMS is to effectively monitor those changes because one of the problems with nonlinear batteries is that their intrinsic characteristics change depending on how many charge-discharge cycles they go through. This examines the creation of a battery performance equation, employs it to determine and predict the SoC and SoH of the battery, and shows how to increase battery protection and performance based entirely on the performance decrease that occurs during battery use. The paper also describes a method for stopping battery intake and spotting faults based entirely on estimated battery performance. The functioning of the advised BMS set of rules is shown in Fig. 5. The battery's performance can be determined by the recommended BMS set of rules, which can gauge the battery's voltage, age, and temperature. When determining a battery's performance, the charge-discharge modern-day is measured inside the battery to determine whether or not the ESS is within the charge-discharge nation [17].

### 3.3 Fuzzy Logic

Firstly, the principle of the fuzzy-based battery management system is to find out the battery efficiency and battery charging-discharging state. When a battery is used, its efficiency declines. A battery's efficiency can only drop when it is in use because it is at its optimum in the starting condition.

#### a. Controller design procedure:

The following steps comprise the fuzzy logic-based battery management system.

1. Recognizing the input and output variables
2. Creating control rules.
3. The creation of a fuzzification method and fuzzy membership functions as well as a method for describing system states in terms of fuzzy sets.
4. Selecting the compositional rule of inference.
5. The de-fuzzification process entails translating the fuzzy control statement into precise control operations

#### b. Fuzzy Logic-based battery management system:

##### 1) Selection of input and output variables

Equation (1) describes a battery's efficiency. By deducting the battery loss  $\eta_{loss}$  from the 100% initial battery efficiency  $\eta_{bat}$ , one can express the efficiency of a battery. Loss can be computed based on the charging and discharging powers, as stated in Equation because the internal resistance increases as efficiency  $\eta_{loss}$  declines (2).

$I_{bat}$ ,  $R$ , and  $V_{bat}$  are the charge-discharge current and internal resistance of the battery, respectively.

$$\eta_{bat} = 100 - \eta_{loss} \quad (1)$$

$$\eta_{loss} = \frac{I_{bat}^2 \times R}{V_{bat} \times I_{bat}} \quad (2)$$

The equation can be used in this situation to express the battery's charge-discharge current (3). The current throughout a battery's charge-discharge cycle can be calculated using the amount of charge and the C-rates at which the battery has been charged or reduced over time. The battery's capacity is denoted by  $Q_{bat}$ , and its charge-discharge cycle is measured by  $t$ . Equation (3) uses the electric charge equation.

The overall decrease in battery efficiency, seen in Eq. (1), is given in Eqs. (2) and (3). (4). The battery efficiency can be determined using Eq. (1), the battery loss equation, and Eq. (4). To figure out a battery's internal resistance, apply Eq. (4). The internal resistance of the battery is provided by Eq. (5). In this study, battery efficiency declined as the charge-discharge process advanced.

$$I_{bat} = \frac{Q_{bat}}{t} \quad (3)$$

$$\eta_{loss} = \frac{\frac{Q_{bat}}{t} \times R}{V_{bat}} \quad (4)$$

$$R = \frac{\eta_{loss} \times V_{bat} \times t}{Q_{bat}} \quad (5)$$

## 2) Selection of Membership Function

Depending on the application, there may be an odd number of linguistic variables used to describe the fuzzy subsets of a variable. Seven is a reasonable number. But as the number of fuzzy subsets rises, the number of rules also rises in direct proportion. There is a fuzzy membership function for each linguistic variable. The membership function converts fuzzy variables from crisp values. The degree of membership, which is utilized to create a fuzzy controller, is defined using the triangle membership functions. Those who are not required to be understood during in the training phase due to their fuzzy-based parameters. As a result, using the following hypothesis, their values are drawn at random from a real-valued function random distribution. This hypothesis will offer the following two decisive rules:

- The rule for  $z(R1 > 40)$ : if the sign is end of battery charge-discharge
- The rule for  $z(R2 > 80)$ : if sign is start discharging
- The rule for  $z(R3 < 20)$ : if the sign is start Charging

## 3) Designing a fuzzy rule:

Utilizing information about the designing system, one can identify a set of rules that specify the relationship between the fuzzy controller's input and output.

### Rule 1:

The voltage determined using the battery's open circuit voltage formula is expressed in Eq. (6). The internal resistance discovered by utilizing using Eq. (5) and the battery's OCV can be used to predict the battery's condition more precisely. So  $H_{after}$  is the SoH of the battery compared to the time after charging and discharging. The SoH of the battery can be predicted using the charge-discharge time of the battery. Here,  $t_{before}$  is the battery charge-discharge time before  $t_{after}$ .

$$t_{after} = \frac{c_n \times (SoC(t) - V_{ocv} - \eta_{loss} \times V_{bat})}{I_{bat}} \quad (6)$$

$$Z(R1) = \frac{t_{after}}{t_{before}} \times 100\% \tag{7}$$

Managing the battery charge-discharge cycle requires the use of SoH → Z (R1), a battery life indicator. To increase the battery’s performance and safety, several methods for predicting SoH have been put forth. The conventional approach analyses a battery according to its chemical principle and models it mathematically or physically to determine how long it will last; however, these approaches do not take into account a battery’s internal resistance, which has a substantial impact on how long it will last.

**Rule 2:**

The equation needs to be rearranged after the SoC → Z (R2 and R3) charge-discharge process using the SoC produced by Eq. (3) and (5). SoC(t) → SoC at time t, SoC(t – 1) → initial SoC, Cn → battery capacity, and V<sub>ocv</sub> → battery voltage in the open state

$$SoC(t - 1) = V_{ocv} + (I_{bat} \times R) \tag{8}$$

$$Z(R1) = SoC(t - 1) + \int_0^t \frac{I(t)}{C_n} dt \tag{9}$$

If the price is inside the over range and the final output values of Eqs. (1) and (9) are detected by the BMS, a rated discharge termination sign is sent over the CAN connection. The look predicted the accurate battery condition using BMS and identified the issue throughout the rate-discharge method to give a BMS set of specifications for an ESS with a significant battery capacity.

**4) Defuzzification**

A fuzzy set is used as the input and a single crisp number is produced after the defuzzification procedure. Although fuzziness aids rule assessment during the intermediate steps, the final desired output for each variable is typically a single number. The centroid calculation is the most widely used defuzzification technique since it yields the center of the area under the curve, which is taken into account during defuzzification.

**4 Result and Discussion**

To test the BMS algorithm of the ESS taking into account battery efficiency-based fuzzy logic, a 3-kW ESS was developed. As several battery modules were charged and discharged, the battery efficiency-based fuzzy logic was put to the test. The ESS was used to implement the BMS algorithm created in this paper. The SoC was subjected to the internal resistance determined by the battery efficiency. The suggested method, OCV, and CCM were then compared, and the SoC was verified in the event of a battery malfunction. By converting the SoC computed from the battery’s internal resistance into the charging-discharging time, the charge-discharge cycle was carried out. Additionally, in the event of a malfunction, the link between the ESS and BMS validated the end of the charge-discharge cycle. The ESS’s overall effectiveness was further proven in the additional algorithmic step to ensure its authenticity.

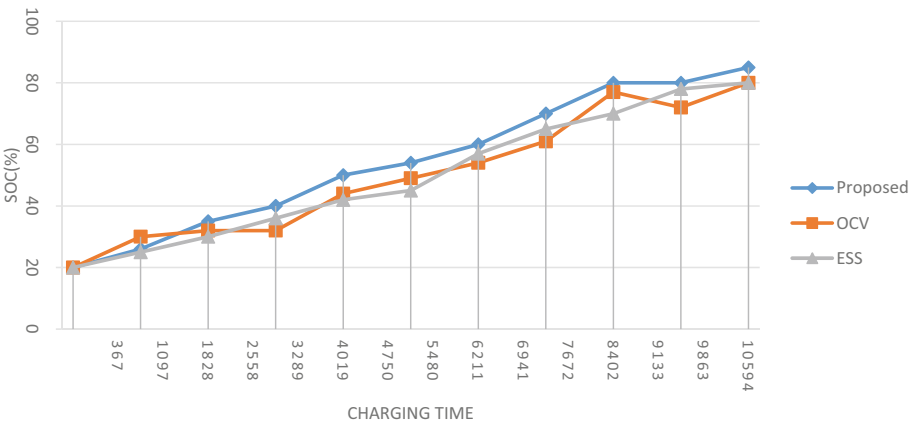
### 4.1 Comparative Analysis

A test battery with three modules was proposed to determine the battery’s efficiency. Modules 1 and 2 were designated as regular batteries, and module 3 was designated as a battery that has undergone numerous charge-discharge cycles.

Each of the three common battery modules has seen a 20% drain and an 80% charge. The OCV, CCM, and suggested SoC set of rules could be compared using battery performance. The application of the suggested set of rules confirms the SoC profile. The OCV and CCM have been compared with the suggested set of SoC calculation rules to validate the SoC calculation. After the preliminary value has been determined, the CCM is charged. The OCV couldn’t because the SoC should be decided the entire charging process. The CCM and the suggested SoC operation seemed to compute the SoC, but when using the actual CCM, the user couldn’t do so immediately and as it should be done by setting the preliminary value. As a result, the SoC may be decided extra as it should be using the set of principles given in this paper. The SoC profile of the battery, to which the suggested battery performance equation is applied, is shown in Fig. 3 and Table 1.

**Table 1.** Comparative analysis of SoC with proposed and existing methods

Algorithm	0	1080	2160	3240	4320	5400	6480	7560	8460	9720	10,800
OCV	20	30	32	32	44	49	54	61	77	72	80
ESS	20	25	30	36	42	45	57	65	70	78	80
Proposed	20	26	35	40	50	54	60	70	80	80	85

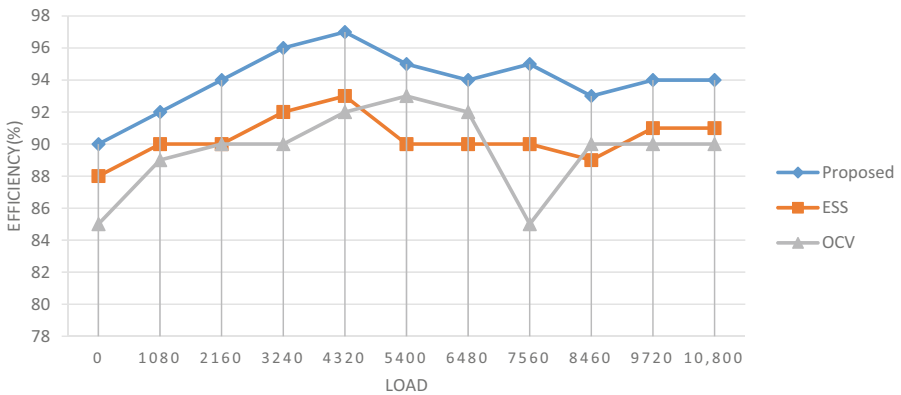


**Fig. 3.** Comparative analysis of SoC with proposed and existing methods

The operating costs associated with utilizing the battery rise as a result of a loss in ESS efficiency, which is produced by a reduction in the difference between the power generated by discharging and the power spent when charging. The effectiveness of the suggested EMS and BMS algorithms was evaluated. The maximum efficiency achieved by the algorithm suggested in this research was 97.57%. The efficiency waveform of the ESS when the system was developed by using the suggested techniques is shown in Fig. 4 and Table 2.

**Table 2.** Comparative analysis of Efficiency with proposed and existing methods

Algorithm	0	1080	2160	3240	4320	5400	6480	7560	8460	9720	10800
Proposed	90	92	94	96	97	95	94	95	93	94	94
ESS	88	90	90	92	93	90	90	90	89	91	91
OCV	85	89	90	90	92	93	92	85	90	90	90



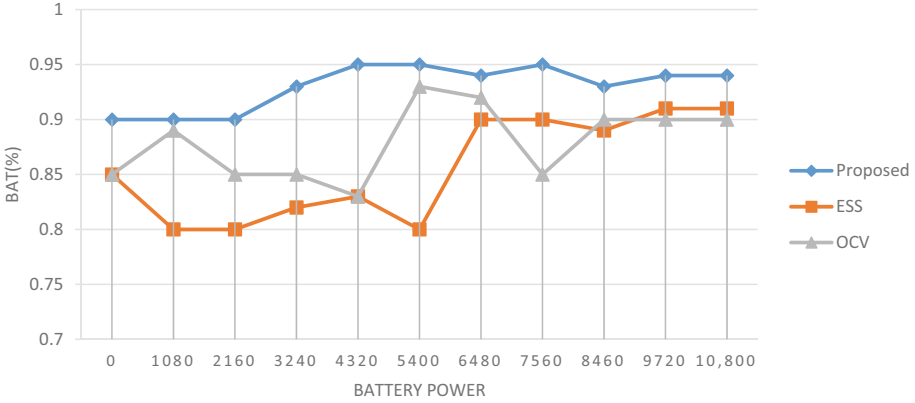
**Fig. 4.** Comparative analysis of Efficiency with proposed and existing methods

The battery performance examination found a significant difference between the performance of the battery and those of the normal battery after comparing the internal changes in the faulty or unusual batteries’ performance that occurred throughout the price-discharge cycle of the ESS. The difference between the performance of old batteries and faulty batteries reached 38.4%. The findings show that it is possible to more accurately predict the battery condition by converting the inner resistance, which has the greatest impact there, as well as to more accurately estimate the SoC and SoH utilization of the suggested set of regulations. Figure 5 and Table 3 show the performance curve of the battery module. As shown in Fig. 4 and Table 3, the difference between the battery’s most recent internal resistance readings was validated during battery charging. If a certain kind was mentioned during the price-discharge cycle for testing, the likely extrude inside the battery’s state of charge due to age can be identified.



**Table 3.** Comparative analysis of Battery Power with proposed and existing methods

Algorithm	0	1080	2160	3240	4320	5400	6480	7560	8460	9720	10800
Proposed	0.9	0.9	0.9	0.93	0.95	0.95	0.94	0.95	0.93	0.94	0.94
ESS	0.85	0.8	0.8	0.82	0.83	0.8	0.9	0.9	0.89	0.91	0.91
OCV	0.85	0.89	0.85	0.85	0.83	0.93	0.92	0.85	0.9	0.9	0.9



**Fig. 5.** Comparative analysis of Battery Power with proposed and existing methods

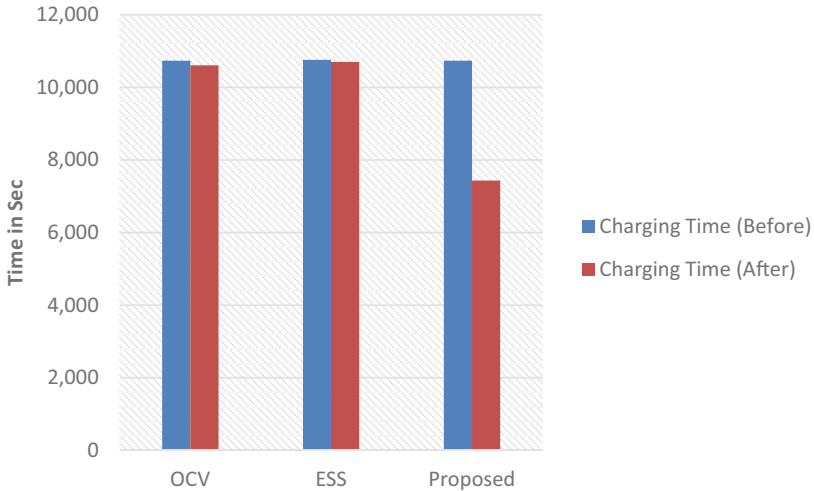
The battery was charged by applying the internal resistance, which changed the battery’s performance. The findings showed that when the battery failed or had other issues, the module’s CC charging time dropped. Equation (10) demonstrated that it was capable of calculating SoH principally based on the battery’s CC termination time by comparing the SoH profiles obtained using the three methodologies investigated. The  $\Delta SoH$  is the amount of change between  $SoH_{before}$  and  $SoH_{after}$ , while  $SoH_{before}$  is the SoH before  $SoH_{after}$ .

$$\Delta SoH = \left( 1 - \frac{SoH_{after}}{SoH_{before}} \right) \times 100 \tag{10}$$

Figure 5 depicts the SoH profile to which the recommended technique was applied, and Table 3 illustrates the CC termination time dependent on battery status (Table 4, Fig. 6).

**Table 4.** The battery efficiency was applied to the SoH table for three battery modules

	Charging Time (Before)	Charging Time (After)	Soc
OCV	10,740	10,610	4.00%
ESS	10,760	10,700	4.00%
Proposed	10,740	7430	35%

**Fig. 6.** Battery efficiency was applied to the SoH profile with CC charging time

This study recommends a BMS technique for an ESS that is based on fuzzy logic. In the experiment, the internal resistance of the battery was determined by deriving its efficiency from the BMS algorithm and applying it to the ESS. Applying battery efficiency to the SoC using the OCV and CCM and the SoH while accounting for charging time, tests were also conducted to show the improvement in battery state accuracy. Additionally, the study backed increased safety by identifying faulty batteries [17].

## 5 Conclusion

In this research, we proposed a fuzzy logic-based BMS algorithm that takes battery efficiency into account. To increase the battery's performance and safety, the algorithm was applied to an ESS. The findings indicate that when the evaluated SOC is incorrect, the BMS is successful in preventing battery over-discharge. In a plug-in series ESS, the fuzzy logic energy-management controller has also shown that it is capable of keeping the engine running in its fuel-economic region. Results from experiments are presented to show how well the suggested fuzzy logic Battery management system performs under various driving circumstances and battery SOC and SoH. The outcomes show that the ESS-based fuzzy logic battery management system was successful in keeping the engine

operating close to its area of optimal fuel efficiency while preventing the battery from charge-discharging. Finally, accurate SoC and SoH estimations were proposed by including battery efficiency in the estimation process. By utilising a defect detection technique with precise SoH estimates, the estimated SoC and SoH were put to use to enhance the performance of the BMS-based fuzzy logic as well as the battery safety.



## References

1. Lipu, M.H., et al.: Intelligent algorithms and control strategies for battery management system in electric vehicles: progress, challenges and future outlook. *J. Clean. Prod.* **292**, 126044 (2021)
2. Xiong, R., Li, L., Tian, J.: Towards a smarter battery management system: a critical review on battery state of health monitoring methods. *J. Power. Sources* **405**, 18–29 (2018)
3. Shen, M., Gao, Q.: A review on battery management system from the modeling efforts to its multiapplication and integration. *Int. J. Energy Res.* **43**(10), 5042–5075 (2019)
4. Li, W., Rentemeister, M., Badeda, J., Jöst, D., Schulte, D., Sauer, D.U.: Digital twin for battery systems: cloud battery management system with online state-of-charge and state-of-health estimation. *J. Energy Storage* **30**, 101557 (2020)
5. Liu, K., Li, K., Peng, Q., Zhang, C.: A brief review on key technologies in the battery management system of electric vehicles. *Front. Mech. Eng.* **14**(1), 47–64 (2019)
6. Xiong, R., Ma, S., Li, H., Sun, F., Li, J.: Toward a safer battery management system: a critical review on diagnosis and prognosis of battery short circuit. *Iscience* **23**(4), 101010 (2020)
7. Ali, M.U., Zafar, A., Nengroo, S.H., Hussain, S., Junaid Alvi, M., Kim, H.J.: Towards a smarter battery management system for electric vehicle applications: a critical review of lithium-ion battery state of charge estimation. *Energies* **12**(3), 446 (2019)
8. Tran, M.K., Panchal, S., Khang, T.D., Panchal, K., Fraser, R., Fowler, M.: Concept review of a cloud-based smart battery management system for lithium-ion batteries: feasibility, logistics, and functionality. *Batteries* **8**(2), 19 (2022)
9. Lin, Q., Wang, J., Xiong, R., Shen, W., He, H.: Towards a smarter battery management system: a critical review on optimal charging methods of lithium ion batteries. *Energy* **183**, 220–234 (2019)
10. Wang, Y., Xu, R., Zhou, C., Kang, X., Chen, Z.: Digital twin and cloud-side-end collaboration for intelligent battery management system. *J. Manuf. Syst.* **62**, 124–134 (2022)
11. Luca, R., Whiteley, M., Neville, T., Shearing, P.R., Brett, D.J.: Comparative study of energy management systems for a hybrid fuel cell electric vehicle-A novel mutative fuzzy logic controller to prolong fuel cell lifetime. *Int. J. Hydrog. Energy* **47**(57), 24042–24058 (2022)
12. Esfandyari, M.J., Yazdi, M.H., Esfahanian, V., Masih-Tehrani, M., Nehzati, H., Shekoofa, O.: A hybrid model predictive and fuzzy logic based control method for state of power estimation of series-connected Lithium-ion batteries in HEVs. *J. Energy Storage* **24**, 100758 (2019)
13. Zheng, W., Xia, B., Wang, W., Lai, Y., Wang, M., Wang, H.: State of charge estimation for power lithium-ion battery using a fuzzy logic sliding mode observer. *Energies* **12**(13), 2491 (2019)
14. Nagaiah, M., Chandra Sekhar, K.: Analysis of fuzzy logic controller based bi-directional DC-DC converter for battery energy management in hybrid solar/wind micro grid system. *Int. J. Electr. Comput. Eng.* **10**(3), 2271 (2020)
15. Panwar, N.G., Singh, S., Garg, A., Gupta, A.K., Gao, L.: Recent advancements in battery management system for li-ion batteries of electric vehicles: future role of digital twin, cyber-physical systems, battery swapping technology, and nondestructive testing. *Energy Technol.* **9**(8), 2000984 (2021)

16. Vivas, F.J., et al.: Multi-objective fuzzy logic-based energy management system for micro grids with battery and hydrogen energy storage system. *Electronics* **9**(7), 1074 (2020)
17. Lee, J., Kim, J.-M., Yi, J., Won, C.-Y.: Battery management system algorithm for energy storage systems considering battery efficiency. *Electronics* **10**(15), 1859 (2021)



# Involvement of Unmanned Aerial Vehicles and Swarm Intelligence in Future Modern Warfare: An Overview

Murat Bakirci  and Muhammed Mirac Özer<sup>(✉)</sup> 

Unmanned/Intelligent Systems Lab, Faculty of Aeronautics and Astronautics, Tarsus University,  
Mersin 33400, Turkey

{muratbakirci, muhammed\_mirac}@tarsus.edu.tr

**Abstract.** This study provides an overview of the potential uses of unmanned aerial vehicles and their swarm systems in future military technology. By examining the usage purposes in different application areas, the basic requirements of a common UAV design are investigated. Critical requirements have been determined, and it has been demonstrated that UAV systems with certain key features can be easily incorporated into swarm systems and used in future military applications. In addition, a common UAV that can be utilized in various applications has been developed, and constraints in the design and production stages have been determined. In this respect, the study provides a basis for similar studies to be carried out.

**Keywords:** Swarm systems · UAV · Intelligence · Military

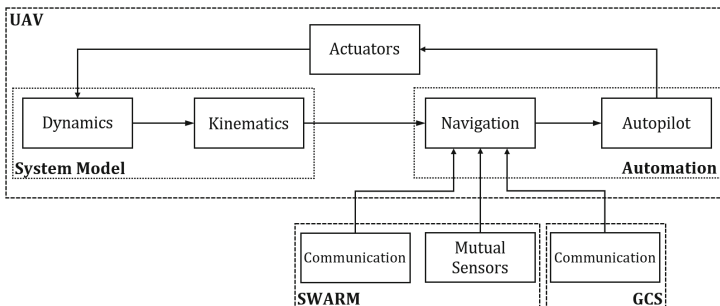
## 1 Introduction

Today, rapidly advancing technology of sensors, processors and communication units enables smaller sized unmanned aerial vehicles (UAVs) to perform the same function as full-scale UAVs. The use of multiple UAVs instead of a single UAV to fulfill a given task has been an area of interest for years [1]. However, there are many scenarios where the capabilities of a single UAV are insufficient due to some constraints such as flight time, payload capacity, and the long-distance communication bandwidth. In recent years, especially after the developments in collision avoidance and path planning, the use of swarm UAVs has increased [2]. The swarm UAV concept refers to UAVs consisting of variety of subsystems that cooperate to perform a specific task.

Basically, swarm UAVs are classified as homogeneous or heterogeneous swarms with advantages and disadvantages depending on sensor characteristics, automation levels, and even platform configurations [3]. In the case of the use of identical UAVs, the structure of the system becomes less complex, and the capabilities of the sensors such as information processing capacity and flexibility remain more limited. Although the group (heterogeneous swarm) of UAVs with different characteristics has a more complex system, it becomes much more useful in conditions where diversity in UAV capabilities

is vital. The aforementioned heterogeneous swarms are used in situations that require a high level of sensor capability or different types of sensors, but sensors of this size cannot be carried by a single UAV due to load capacity constraints [4].

Comprehensive studies in the literature on the control (route planning and task sharing) of multiple UAV systems are mainly based on centralized [5] and decentralized/decentralized [6] architectures. In a centralized control approach, where there is a low level of autonomy and no mutual communication between the UAVs, the operator(s) receives information from each UAV system, coordinates it, and pre-defines the task assignments for each UAV system. While this approach is simpler and easier to optimize due to its low level of autonomy, it has less redundancy and is less resilient to failures in the UAV or communications. In the second approach, the decentralized (distributed) control approach, there is a high level of autonomy in the UAV system and mutual communication between UAVs and ground control station (GCS) as shown in Fig. 1. The role of the operator is that each UAV should be able to communicate, receive and share information and make the necessary decisions so that it can reach a higher managerial level. Decentralized systems could not be used until recent years because they are more complex and require a high level of autonomy. The importance of the decentralized approach, as the mission and information are distributed among the UAVs, the system is more durable and flexible and includes more redundancy. In addition, since the holistic behavior of the system is built on local cooperation between UAVs, UAVs adapt more easily to the dynamic environment. In addition, since this approach is based on system cooperation, it is more resistant to temporary or permanent errors that occur in individual UAVs.



**Fig. 1.** Control scheme of a member UAV of the drone swarm.

Drone swarms are deployed in three types as static swarm, dynamic swarm and hybrid swarm, considering the mission environment and situation [7]. The first and most basic type of swarm is that the members of the static swarm are pre-selected at the pre-assignment stage and new members cannot enter the swarm during the flight as they are prevented from joining the group at the starting point of the mission. Secure communication, mutual trust and cooperation are created by the flight control system on the ground. During the flight, each UAV belongs to the swarm. If the UAV does not belong to the swarm, it is treated as an entity outside the swarm. In contrast to static swarm, the second type, dynamic swarm, allows members to leave the swarm

at any time (before/during the mission) and allow new members to be included in the swarm. A closed-dynamic swarm only allows new drones from the same organization to join the swarm, while an open-dynamic swarm allows drones from any supporting (third party) organization to join. Both situations involve challenges in establishing secure communication, mutual trust, and collaboration. The last type, hybrid swarm type, combines both static and dynamic swarms into a single working whole. At the core of this swarm is a static swarm that acts as a single swarm in all operations. This static swarm allows other UAVs to join, creating an expanded swarm that acts as a dynamic swarm. In addition, the center swarm is a high priority when it comes to any collaborative learning, assessment and decision making. In the extended swarm, UAVs joining the core (static) swarm can be considered to provide service to this swarm in exchange for a fair exchange. Members of the extended pack can leave the group at any time.

Considering communication models between UAVs in order to analyze and optimize UAV communication strategies, the key to problem solving in multi-element swarms is to provide multi-coordination. It is important that the members of the swarm communicate in a proper coordination in order to reach the information needed to decide which actions to take. Since UAVs move quickly and flexibly, it is not difficult for them to communicate over wire structures. Considering their flight behavior and speed, it is more appropriate for UAVs to use radio devices. Therefore, communication, while many control systems are being developed, careful thought is taken at the element level. In terms of radio signal coverage, there are two types of communication, Direct and Indirect Communication, in terms of signal delay, coverage area and consumed power in communication between swarm elements [8].

In the first of these, the direct communication, UAVs communicate directly from a single center in a limited coverage area, without a base station, using channels such as WiFi, Bluetooth and ZigBee. Their advantage is low power consumption, less latency and low cost. On the other hand, the indirect communication, which is the other type, UAVs communicate via ground-supported base stations or supported base stations with moving vehicles without distance restrictions. As the coverage area expands in long distance indirect communication (such as GSM and satellite), the power consumption, cost and delay increase.

At the application level, there are three different types of communication: Broadcast, Synchronization and Query. During the broadcast, a group member shares its current status with the other members of the group. During the query, the member sends a message to the other members of the group asking them to report their current status. During synchronization, the group leader sends a message containing a sync request to the other members of the group, and all members receiving this request broadcast their current status to other members in the group.

In the light of all this, the control mechanism designer can determine the design principles by choosing the direct or indirect communication tools appropriately to transmit the messages. For example, in order to learn the status of the UAVs within the coverage area of the direct communication device, only one query message will be needed to be sent to all UAVs on the direct communication path, and the UAVs currently within the direct communication coverage area will also receive this message and report their status.

## 2 Collaborative Swarm Systems

Swarm UAV systems can be considered as a subset of swarm systems in terms of their features/capabilities. Accordingly, multi-UAV systems have some disadvantages compared to single-UAV systems. Considering its advantages, it offers improved performance by segregating and parallelizing tasks and performing these tasks more effectively as a group. However, some missions that cannot be performed by a single UAV, as in natural systems, can be achieved by swarm UAVs and provide “mission activation”. In addition, the information gathered by the swarm UAVs by forming a “sensor network” will be much more comprehensive than the information obtained from the coverage area of a single UAV, thus providing “distributed detection”. In addition, “distributed action” is provided by carrying out many simultaneous joint actions in different places at the same time. The last point to be mentioned is that since the failure of a single UAV in the swarm does not prevent the fulfillment of the given task, the “fault tolerance” decreases.

Although these systems have advantages, they also bring some disadvantages like any other systems. Considering these disadvantages, firstly, the problem of interference arises. In this context, the actions of individual UAVs in a group (even those that are coordinated) can be mutually influenced by each other (collisions, loss of communication, etc.). Associated with this, due to uncertainty in the intentions of other UAVs, requires each UAV to fully understand what it is doing with other UAVs and what it expects nearby robotic platforms to do. In the case of uncertainty (or when the threshold level is lower than clarity), UAVs can compete rather than cooperate. The last issue is the overall system cost. Here, too, single UAV systems are typically more “complex” and also more expensive, while individually less complex UAV swarms are less costly than single UAV systems.

A swarm can perform many military and civilian tasks by flying in a certain order. For example, in exploration work in a specific area, tracking forest fires and fighting forest fires [9], monitoring and controlling buildings, bridges and hard-to-reach places [10], control, surveillance of a private or military area for unauthorized access [11], aerial photography, search and rescue operations after natural disasters such as floods and earthquakes, and re-establishing communication lost during disasters [12], agricultural spraying, tracking the development of crops in the field etc., can be used.

Three different use cases for systems with swarm intelligence can be discussed. The first of these is Remote Rescue Operations [13]. The swarm can be used in different types of rescue operations in remote areas, for example, when an event such as an earthquake makes it difficult and dangerous for emergency teams to access a certain area. The swarm also enables the establishment of a communication network for the information exchange of the emergency teams or the improvement of the network in general use (3G/4G) to communicate with the emergency teams or families of the people in the disaster area. UAVs can quickly provide situational awareness over a wide area, significantly reducing the cost and risks of search and rescue efforts by reducing the number and time of employees in finding and rescuing the injured or missing. It also enables it to contribute to the prioritization of search and rescue efforts by displaying the damage caused by the disaster. It helps ensure public safety in a wide variety of ways. In addition, it can be used in rescue operations such as searching for signs of life in an avalanche area or searching/scanning the survivors of a shipwreck on the sea. In such events, larger



areas can be covered with swarm UAVs compared to conventional vehicles (helicopter, airplane or ship, etc.).

The second use case is Plant Monitoring and Fault Detection [14]. UAVs can be used for surveillance of events such as fire detection in wide area parks and forests with multiple sensors (thermal camera, day camera, etc.). It also offers an additional level of security against intrusion or disruption in highly secure buildings or large infrastructure facilities. When used for border security, they provide cost-effectiveness by eliminating the need for manned patrols or wall construction.

The last area of use considered is Data Collection [15]. When a disaster or event threatens human life and livelihoods, emergency responders need information and real-time images to make better decisions and save time. The most basic and most common use case for data collection in such cases involves a swarm of drones collecting ground images for various purposes. For example, observing disasters (nuclear accidents, building fires, ship collisions, plane and train accidents, motor vehicle accidents, terrorist activities), observing natural disasters (landslides, earthquakes and tsunamis, fires, volcanic eruptions, floods, storms, hurricanes, avalanches and tornadoes), monitoring military targets (identifying the enemy's position), and monitoring agricultural areas (identifying areas in need of irrigation or care for a spreading disease).

In recent years, UAVs have been attracting more and more attention in both the civilian and military fields, thanks to the reduced deployment and maintenance costs of UAV systems, and their ability to operate in areas that are inaccessible and dangerous for pilots. For example, the use of UAV swarms can be envisaged to tackle complex tasks such as searching for survival cues, multi-target tracking and tracking. These swarms require more advanced control, communication and coordination mechanisms. However, it becomes difficult to test and analyze these mechanisms in dynamic flight conditions.

### 3 Exclusive UAV Swarm Systems

The use of UAVs in disaster management and relief operations is a topic currently being studied and discussed in various research projects. In this context, the German Federal Ministry of Education and Research (BMBF) developed the AirShield (Airborne Remote Sensing for Hazard Inspection by Network Enabled Lightweight Drones) project in July 2008 for a system that collects data from the air about a dangerous situation using autonomous mobile aerial robots with small and light sensors [16]. The system supports decision-making abilities of public authorities and other organizations by identifying a fire in a rural area and estimating the threat that this fire may cause, by presenting the information collected from the air to public authorities and other organizations. Besides Germany, countries such as US, UK, and South Korea are also interested in development of unmanned reconnaissance systems for disaster management.

A technology was developed by a consortium under the coordination of Cranfield University to enable heterogeneous UAVs to work as a swarm in the military environment, and integration was carried out within the scope of this project called EuroSWARM. The system was tested and verified using the information obtained from simulation studies and small-scale experiments. The mission of the study is to perform information fusion and to show that swarm technology can be developed by assigning

tasks in a mobile structure. With the cooperation of consortium partners, demonstration scenarios, system/subsystem requirements and performance metrics for the solutions to be developed in the project are defined. The technical review has shown that it will be challenging to achieve the proposed innovations. With this, a feasibility study conducted at the initial stage shows that these innovations can be realized with the proposed approach and that this will enable real-time and situational awareness information to be obtained before contact is made in the military field, providing information that will help commanders in making appropriate decisions, and reducing the risk of mission and operation [17].

With the SAGA Project (Swarm Robotics for Agricultural Applications), which was developed for agricultural applications, it is aimed to monitor the development of agricultural products, to determine the formation of weeds, mapping, and also to weed them, to reduce the cost and to increase the amount and quality of the product with the robot herds spread over the agricultural field [18]. A widely attended European Union project called CPSwarm aims to perform tasks such as search and rescue and surveillance of critical infrastructures such as industrial facilities and power plants with heterogeneous swarms of ground robots/mobile vehicles and UAVs [19].

In October 2018, Airbus successfully carried out test flight campaigns for manned-unmanned crewing for future air combat systems, demonstrating its ability to control unmanned systems from manned aircraft using five Airbus-made Do-DT25 target drones. During flight tests at a test site in the Baltic Sea region of Germany, the drones were controlled by the task group commander on the manned command and control aircraft. In test flights, elements such as inter-system connection, human-machine interface and concepts such as team intelligence were verified. Unmanned systems equipped with sensors provide situational awareness to a task group commander at a safe distance in a manned aircraft. The expertise gained in these studies is used to develop the Future Air Combat System for Europe [20]. China, on the other hand, has been rapidly developing the use of UAVs in recent years, collecting highly reliable and comprehensive data, while saving time and money. Moreover, multiple UAVs operating in groups offer an opportunity for a new operational paradigm.

## 4 A UAV Use Case

Considering the way the aircraft are commanded, since it is technically easier to make route planning in divided space and assign position reference points defined in the flight software to certain coordinated points, route planning algorithms suitable for this situation have been evaluated and have been able to find the most application area among them for many years and have been mathematically analyzed. The A\* algorithm, whose behavior can be considered reliable, is quite popular. The most important benefit of this algorithm is that it can be easily modified according to different requirements. Although planning in continuous space can create more sophisticated and flexible routes, planning for more than one unit at the same time, this planning for relatively large areas for real flights, and RRT (Rapidly-Exploring Random Tree), PRM (Probabilistic Roadmap), and derivatives algorithms are generally not preferred. The main reason for this is that these algorithms create sampling points in continuous space and do route planning by connecting these points, but this sampling and creating a connected network between them

is computationally heavy. As mentioned above, planning for many members simultaneously complicates the situation or increases the required planning time. On the other hand, since these algorithms do not guarantee to find any route mathematically, they can cause problems on the application. Finally, thanks to increased computational capacities, computers can now create routes directly defined by kinematic reference points for aircraft, rather than generating waypoints. In short, it is the creation of reference points for the required position, speed, acceleration and its numbered derivatives on different axes for every moment the aircraft is in, and the aircraft follows it. However, this option is not applicable because it creates many security vulnerabilities to perform this operation on autonomous control on the autopilot used.

Moreover, the software environment and communication infrastructure in which they are applied are as important as the success of motion algorithms. At this point, the use of Robot Operating System (ROS) infrastructure, which has been used frequently in autonomous robotic applications and provides a systematic communication solution for communication between sensors, controllers or members, is appropriate for an autonomous system with multiple members. Through a detailed literature review, Gazebo simulation environment, which is a preferred simulation infrastructure together with ROS, stands out. Offering easy-to-use models for many different robotic platforms, widespread use, and flexibility in working with ROS make Gazebo stand out compared to alternatives such as VREP, AirSim and X-Plane.

The platform developed for common purposes, the CAD design of which is shown in Fig. 2, is a custom-built rotary-wing drone with a weight of 355 g, a width of 455 mm from rotor to rotor and a height of 270 mm. In addition to the MPU-9250 3-axis gyroscope, accelerometer and magnetometer, there is also an LPS25H pressure sensor to assist in altitude estimation. The recommended maximum overhead is determined as 105 g, while the custom-built system can stay airborne for about 22 min. Although it runs all low-level control algorithms on the vehicle, it must receive commands from a computer in order to perform autonomous or pilot-controlled flight. This requires the platform to be in connection with the ground control station (GCS) for all flight scenarios, including autonomous mission scenarios. It provides its connection with the GCS via radio signals, with an apparatus to be attached to the computer. All this is accomplished with a dual MCU structure, the nRF51822 radio and power management microcontroller (Cortex-M0, 32 MHz, 16kb SRAM, 128kb flash), as well as the main application microcontroller (Cortex-M4, 168 MHz, 192kb SRAM, 1Mb flash) powered by STM32F405. While the platform is being designed, it offers powerful options such as logging, real-time parameter adjustment, and wireless firmware updating, so that it can easily develop on the product with low latency and long range radio or BLE communication. In addition, it enables to do all these with a smart phone or tablet, using it as a remote. On the other hand, for many software projects, the product IDE provides virtual machines, eliminating the need to install some software. The developed prototype of the designed rotary wing system is shown in Fig. 3.

In such systems, access to global positioning systems may not be provided in a proper way for several constraints. Thus, an external position feedback is needed. For this purpose, Prime 13 model motion capture cameras of the Optitrack brand, which carry filters





**Fig. 3.** Custom-built rotary wing UAV.

## 5 Conclusion

Unmanned aerial vehicles and swarm systems formed by their multiple and coordinated use are potential candidates that can be developed to higher levels. It is aimed to make a difference in this developing area and to enable unmanned aerial vehicles to perform tasks in different places with swarm intelligence. In this context, an extensive literature survey was conducted, and system requirements for a common UAV that can be used in almost all applications were determined by making an evaluation based on the current usage types. Thus, the materials, external equipment, types of software and interface, flight mode and usage instructions required for the proposed project were determined. It can be deduced that it is highly critical to know whether the design parameters determined in line with the system requirements are applicable with the available resources, in terms of the implementation of such a research and development study. As in similar exclusive examples, the necessity of a coordinated and collaborative work has emerged. Otherwise, there may be a return from the production phase to the initial design, and many fundamental changes may be needed, which can cause time and resource loss.

## References

1. Wei, N., Jiaqin, H., Lifeng, M.: Research on the concept and key technology of UAV swarm combat against the sea. *Command Control Simul.* **40**(1), 20–27 (2018)
2. Yu, Y., Zhang, Y., Jiang, B., Fu, J., Jin, Y.: A review on fault-tolerant cooperative control of multiple unmanned aerial vehicles. *Chin. J. Aeronaut.* **35**(1), 1–18 (2022)
3. Wu, M., Zhu, X., Ma, L., Wang, J., Bao, W., et al.: Torch: strategy evolution in swarm robots using heterogeneous–homogeneous coevolution method. *J. Ind. Inf. Integr.* **25**, 100239 (2022)
4. Du, W., Ying, W., Yang, P., Cao, X., Yan, G., et al.: Network-based heterogeneous particle swarm optimization and its application in UAV communication coverage. *IEEE Trans. Emerg. Top. Comput. Intell.* **4**(3), 312–323 (2019)
5. Indriyanto, T., Rizki, A.R., Hariyadin, M.L., Akbar, M.F., Syafi, A.A.A.: Centralized swarming UAV using ROS for collaborative missions. In: *AIP Conference Proceedings*, vol. 2226, no. 1, p. 030012 (2020)

6. Xiao, Y., Ye, Y., Huang, S., Hao, L., Ma, Z., et al.: Fully decentralized federated learning-based on-board mission for UAV swarm system. *IEEE Commun. Lett.* **25**(10), 3296–3300 (2021)
7. Yu, Z., Si, Z., Li, X., Wang, D., Song, H.: A novel hybrid particle swarm optimization algorithm for path planning of UAVs. *IEEE Internet Things J.* **9**(22), 22547–22558 (2022)
8. Campion, M., Ranganathan, P., Faruque, S.: UAV swarm communication and control architectures: a review. *J. Unmanned Veh. Syst.* **7**(2), 93–106 (2018)
9. Madridano, Á., Al-Kaff, A., Flores, P., Martín, D., de la Escalera, A.: Software architecture for autonomous and coordinated navigation of UAV swarms in forest and urban firefighting. *Appl. Sci.* **11**(3), 1258 (2021)
10. Liu, J., Liao, X., Ye, H., Yue, H., Wang, Y., et al.: UAV swarm scheduling method for remote sensing observations during emergency scenarios. *Remote Sens.* **14**(6), 1406 (2022)
11. Wang, Y., Bai, P., Liang, X., Wang, W., Zhang, J., et al.: Reconnaissance mission conducted by UAV swarms based on distributed PSO path planning algorithms. *IEEE Access* **7**, 105086–105099 (2019)
12. Din, A., Jabeen, M., Zia, K., Khalid, A., Saini, D.K.: Behavior-based swarm robotic search and rescue using fuzzy controller. *Comput. Electr. Eng.* **70**, 53–65 (2018)
13. Horyna, J., Baca, T., Walter, V., Albani, D., Hert, D., et al.: Decentralized swarms of unmanned aerial vehicles for search and rescue operations without explicit communication. *Auton. Robot.* **47**(1), 77–93 (2023)
14. Stolfi, D.H., Brust, M.R., Danoy, G., Bouvry, P.: CONSOLE: intruder detection using a UAV swarm and security rings. *Swarm Intell.* **15**(3), 205–235 (2021)
15. Hildmann, H., Kovacs, E., Saffre, F., Isakovic, A.F.: Nature-inspired drone swarming for real-time aerial data-collection under dynamic operational constraints. *Drones* **3**(3), 71 (2019)
16. Daniel, K., Dusza, B., Lewandowski, A., Wietfeld, C.: AirShield: a system-of-systems MUAV remote sensing architecture for disaster response. In: 3rd Annual IEEE Systems Conference, Vancouver, BC, Canada, 2009, pp. 196–200 (2009)
17. Gibney, E.: Europe’s controversial plans to expand defence research, vol. 569, pp. 476–477. Springer, Nature (2019)
18. Shahzad, M., Saeed, Z., Akhtar, A., Munawar, H., Yousaf, M.H., et al.: A review of swarm robotics in a nutshell. *Drones* **7**(4), 269 (2023)
19. Kaur, M., Razi, A., Cheng, L., Amin, R., Martin, J.: Design and evaluation of an application-oriented data-centric communication framework for emerging cyber-physical systems. In: 20<sup>th</sup> Consumer Communications Networking Conference, Las Vegas, NV, USA, 2023, pp. 1–6 (2023)
20. Heilemann, F., Lindner, S., Schulte, A.: Experimental evaluation of tasking and teaming design patterns for human delegation of unmanned vehicles. *Hum.-Intell. Syst. Integr.* **3**(3), 223–240 (2021)



# A Comparative Study of Deep Learning Loss Functions: A Polyp Segmentation Case Study

Rachid Bourday<sup>(✉)</sup>, Issam Aattouchi, and Mounir Ait Kerroum

Computer Science Research Laboratory, Faculty of Sciences, Ibn Tofail University, Kenitra,  
Morocco

Rachid.bourday@uit.ac.ma

**Abstract.** Colorectal cancer is the third most common cancer diagnosed worldwide. The early detection of this disease can help in treating cases and save human lives. Deep learning algorithms appear as interesting tools used to successfully detect and segment polyps, thus improving surgical resection. When implementing these algorithms, there are different architectures to use. Depending on the architecture used, the obtained accuracy and the convergence speed of the algorithms can be different. To assess how these algorithms model datasets, we can use different metrics. In this paper, we compare different segmentation loss function metrics using one conventional architecture called U-Net. The performance was evaluated on three well-known polyp datasets, namely CVC-ClinicDB, Kvasir, and ETIS-Larib PolypDB. Findings show that the best results from CVC-ClinicDB are 85.89% for Dice, 84.89% from Kvasir, and 77.02% from ETIS-Larib PolypDB. The model behaves well even if we combine the three datasets. In fact, the accuracy level still reached 76.71%.

## 1 Introduction

In Morocco, colorectal cancer (CRC) is the third most common cancer among men and women. Published statistics mention that in 2020, 4558 new cases of CRC occurred [1]. Most reported CRC cases are diagnosed at advanced stages. Polyps, noncancerous growths that often develop on the inner wall of the colon or rectum as people age, serve as the starting point for colorectal cancer. If left untreated or unremoved, polyps can progress into dangerous cancers. There are multiple types of polyps, which can be detected through a colonoscopy. A colonoscopy is a procedure used to identify and evaluate polyps by performing biopsies and removing them. Early detection of the disease greatly impacts survival rates in colorectal cancer, making the detection of polyps crucial. Furthermore, studies have shown that polyps are frequently missed during colonoscopies. In fact, approximately 10% of colon polyps are flat and difficult to detect without the use of dyes to highlight them [2]. Additionally, the rate of missed polyps ranges from 14% to 30%, depending on the type and size of the polyp. This can be attributed to the skill of the endoscopist, as colonoscopy is a procedure dependent on the operator's proficiency. Polyp detection is not the only critical stage; segmentation is also challenging due to variations in the shape and color intensity of polyps in colonoscopy frames. With advancements in hardware and algorithmic breakthroughs like deep learning, it is now possible to build accurate real-time systems for polyp detection and segmentation.

In this work, we utilize a CNN deep learning algorithm for polyp segmentation, adapting the U-Net architecture [3], which is renowned for its extensive usage in medical imaging segmentation. By training models with various proposed loss functions (BCE loss, Dice loss, Jaccard loss, Tversky loss, and Focal Tversky loss), we have successfully identified the optimal model that achieves significantly better Dice scores and IOU, resulting in improved predictions. To evaluate our method, we employ well-known publicly available datasets: Kvasir [4], CVC-clinicDB [5], and ETIS-LaribPolypDB [6]. Furthermore, we combine these datasets to assess the performance of our model.

All steps are implemented using a code that we developed under the Python programming language. The remainder of the paper is organized into five sections. The first section introduces loss functions and metrics. Section 3 presents the datasets and their processing. Section 4 provides detailed information about the U-Net architecture and model fitting. Section 5 presents the results obtained from the different datasets used, along with a comprehensive discussion. Finally, Sect. 6 concludes the paper.

## 2 Loss Functions and Metrics

### 2.1 Loss Function

The success of a model is strongly influenced by the objective function, also known as the loss function. This function plays a vital role as it maps crucial information from the data to a measurable cost space, enabling the estimation of costs through optimization algorithms such as stochastic gradient descent [7]. In the domain of semantic segmentation, there are several types of objective functions available. In this study, we introduce the five loss functions that were employed. Binary Cross-Entropy (BCE):

Cross-Entropy is an approach that facilitates the mapping of data information into a categorical distribution. Specifically, Binary Cross-Entropy [8] is employed when dealing with two categorical classes. The distribution is defined using the following formula:

$$L_{BCE}(Y, \hat{Y}) = \frac{1}{M \cdot N} \sum_{i=1}^{M \cdot N} (-y_i \cdot \log(\hat{y}_i) - (1 - y_i) \cdot \log(1 - \hat{y}_i))$$

Let  $X$  the image to be segmented, such as each pixel  $x$  of  $X$  has a corresponding value,  $y$  in the ground truth mask  $Y$ . The size of the image is  $M \times N$ . For this binary segmentation problems, the ground truth labels can take values  $y = \{0, 1\}$ , where 1 is the positive class, in our case the polyp, and 0 is the negative class, corresponding to the background class. The segmentation model gives a prediction  $\hat{Y}$ , labeled  $\hat{y} = \{0, 1\}$ .

In the semantic segmentation context, the minimization of the loss function is used to reach the minimum error for the prediction of the binary mask, considering both the positive class, or polyp, and the negative class, or background.

#### 2.1.1 Dice Loss

The Dice coefficient [9] serves as a measure of similarity between two regions and is equivalent to the F1-score.



The Dice loss can be defined as follows:

$$L_{\text{Dice}}(Y, \hat{Y}) = 1 - \frac{2 \cdot |Y \cap \hat{Y}|}{|Y| + |\hat{Y}|} = 1 - \frac{2 \sum_{i=1}^{M \cdot N} y_i \cdot \hat{y}_i}{\sum_{i=1}^{M \cdot N} y_i + \sum_{i=1}^{M \cdot N} \hat{y}_i} = 1 - \frac{2 \cdot TP}{2 \cdot TP + FN + FP}$$

where:

- $Y$  and  $\hat{Y}$  are defined above ( see Binary Cross-Entropy (BCE) metric)
- True positives (TP): These represent the polyp pixels present in the ground truth binary mask that are correctly classified as polyp in the predicted binary mask.
- False negatives (FN): These refer to the polyp pixels in the ground truth binary mask that are incorrectly classified as background in the predicted binary mask.
- False positives (FP): These indicate the background pixels in the ground truth binary mask that are incorrectly classified as polyp in the predicted binary mask.
- True negatives (TN): These represent the background pixels in the ground truth binary mask that are correctly classified as background in the predicted binary mask.

### 2.1.2 Jaccard Loss

The Jaccard index [10], also referred to as the intersection over union (IoU), calculates the ratio of the intersection between the ground truth and predicted binary masks to their union. It yields a value between 0 and 1, where 0 indicates no overlap and 1 represents a perfect match. When there is uncertainty in the solution, the convention is to consider 0/0 as 1. The Jaccard loss can be defined as follows:

$$L_{\text{Jaccard}}(Y, \hat{Y}) = 1 - \frac{|Y \cap \hat{Y}|}{|Y \cup \hat{Y}|} = 1 - \frac{\sum_{i=1}^{M \cdot N} y_i \cdot \hat{y}_i}{\sum_{i=1}^{M \cdot N} y_i + \sum_{i=1}^{M \cdot N} \hat{y}_i - \sum_{i=1}^{M \cdot N} y_i \cdot \hat{y}_i} = 1 - \frac{TP}{TP + FN + FP}$$

### 2.1.3 Tversky Loss

The Tversky index [11] offers the ability to adjust the penalties for false positives (FP) and false negatives (FN) by utilizing the parameters  $\alpha$  and  $\beta$ . This index is defined as follows:

$$TI = \frac{TP}{TP + \alpha FN + \beta FP}$$

where  $\alpha + \beta = 1$ .

### 2.1.4 Focal Tversky Loss [12]

Based on the Tversky index and combining the idea of the binary focal loss and it is defined as:

$$L_{\text{FocalTversky}}(\mathbf{Y}, \hat{\mathbf{Y}}) = \left(1 - TI(\mathbf{Y}, \hat{\mathbf{Y}})\right)^{\frac{1}{\gamma}}$$

$\gamma$  is a parameter that smoothly adjusts the rate at which easy examples are down-weighted.

## 2.2 Metrics

To assess the performance of our model, we utilize the Intersection over Union (IoU) and Dice coefficient metrics. These metrics evaluate the predicted mask against the ground truth mask from the test dataset. The formula for calculating the IoU is defined as follows:

$$IOU = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

In this formula, the Area of overlap is the common area of two predict masks while the Area of Union represents all area of the two masks.

The Dice Coefficient [13] has the following formula:

$$\text{Dice score} = \frac{2 \cdot |A \cap B|}{2 \cdot |A \cap B| + |B \setminus A| + |A \setminus B|} = \frac{2 \cdot |A \cap B|}{|A| + |B|}$$

This formula calculates the ratio of the intersection area between two masks to the union area of the same two masks.

## 3 Dataset and Processing

### 3.1 Dataset

In this study, we utilized four different polyp image datasets: CVC-ClinicDB, Kvasir-SEG, ETIS-LaribPolypDB, and a combined dataset.

The Kvasir-SEG dataset, introduced by Jha et al. (2020) [4] and Borgli et al. (2020) [14], consists of 1000 polyp images along with their corresponding ground truth. The images in this dataset have varying resolutions, ranging from  $332 \times 487$  to  $1920 \times 1072$  pixels.

CVC-ClinicDB is a database of colonoscopy video frames that has been extensively used in studies by Bernal et al. (2015) [5] and Fernandez-Esparrach et al. (2016) [15]. This dataset contains frames showcasing different instances of polyps. Each polyp is accompanied by a ground truth annotation in the form of a mask that outlines its region in the image. CVC-ClinicDB was the designated training database for the MICCAI 2015 Sub-Challenge on Automatic Polyp Detection in Colonoscopy Videos. It comprises 612 sequential WL (white-light) images featuring polyps extracted from 31 different sequences, representing 31 distinct polyps acquired from 23 patients.

ETIS-LaribPolypDB is a dataset specifically designed for the polyp detection sub-challenge of the ‘Endovis’ Challenge, as described by Silva et al. (2014) [6]. This dataset consists of 196 white-light (WL) images depicting polyps. These images were extracted from 34 different sequences, representing 44 unique polyps. Additional statistical information about the dataset can be found in Table 1, as presented in the work by Lou et al. (2022) [16].

**Table 1.** Public datasets statistics

Dataset	Image size	Number of images	Polyps size
Kvasir-SEG	1070 × 1348	1000	0.79%–62.13%
CVC-ClinicDB	288 × 384	612	0.34%–45.88%
ETIS-LaribPolypDB	966 × 1225	196	0.11%–29.05%

### 3.2 Data Splitting

To evaluate the performance of the machine learning algorithms, we randomly divided our datasets into two separate parts: a training subset comprising 80% of the dataset, and a test subset comprising the remaining 20%. We employed the Adam optimizer to minimize our loss functions, with a learning rate set to  $10^{-3}$ . For the Kvasir-SEG, CVC-ClinicDB, and the combined dataset, we trained our network for 50 epochs, implementing early stopping with a patience of 5. The batch size used for these datasets was set to 8. As for the ETIS-LaribPolypDB dataset, we trained the network for 100 epochs, and the batch size was set to 6.

By employing these training configurations, we aimed to optimize the performance of our network and achieve accurate results for polyp segmentation.

### 3.3 Data Augmentation

Training models for polyp segmentation can be challenging due to the limited availability of training data and the variations in endoscopy images. Endoscopy procedures involve camera movement, leading to inconsistent color calibration and significant differences in image appearance across different laboratories. To overcome this challenge, data augmentation techniques are employed to expand the training data and capture the various variations in polyp images. By augmenting the training data, we can mitigate overfitting issues during model training.

In this study, we applied several data augmentation techniques to enhance the training dataset. These techniques include Horizontal Flip, Vertical Flip, Optical Distortion, Center Crop, and Random Rotate90. Each original image and its corresponding mask were augmented to generate five new artificial images and masks. As a result, each original image contributed to a total of six images and masks (one original and five artificially generated). Figure 1 provides examples of the data augmentation process applied to the original polyp image, illustrating how these techniques can introduce variations and enrich the training dataset.

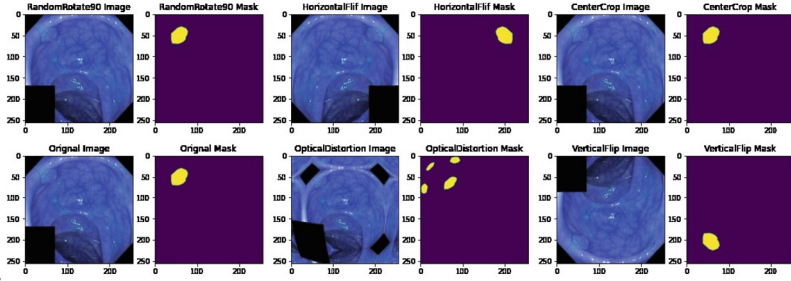


Fig. 1. Polyps and corresponding masks in data augmentation

### 3.4 Data Normalization

Image normalization is a widely used technique in image processing that aims to adjust the range of pixel intensity values. The goal is to convert the input image into a range of pixel values that are more standardized and aligned with human perception, hence the term “normalization.” In our study, we have developed a function specifically designed to perform normalization on both the input RGB image and its corresponding grayscale format.

The normalization process we employ follows a linear approach, utilizing the following formula:

$$Y = ((X - \min) * 255 / (\max - \min))$$

where  $X$  is the pixel value of the original image,  $Y$  is the Pixel value of output image (after normalization),  $\min$  is the minimum pixel value for the original image and  $\max$  is the maximum pixel value for the original image.

## 4 Model Fitting

Once the data has been prepared, the model is fitted using the U-Net network architecture, as illustrated in Fig. 2. The U-Net architecture consists of an encoder (downsampler) and a decoder (upsampler) connected by a bottleneck. In the image, the gray arrows represent the skip connections that link the outputs of the encoder blocks to each corresponding decoder stage.

To enhance modularity and code organization, it is recommended to create functions for the encoder blocks, as they will be repeated throughout the architecture. Each encoder block consists of two Conv2D layers with ReLU activation, followed by a MaxPooling layer and, optionally, a Dropout layer. With each block, the number of filters increases while the dimensions decrease due to the pooling layer. In an encoder block, the height and width of the image decrease while the number of filters increases. Following the encoder blocks, the bottleneck layer extracts additional features. This section does not include a pooling layer, resulting in unchanged dimensions.

Finally, the decoder resamples or upsamples the features to restore the original image size. At each upsampling level, the output of the corresponding encoder block is concatenated and fed into the subsequent decoder block. In a decoder block, the height and width of the image increase while the number of filters decreases.

For the training experiments, the size of the input images and their corresponding masks in the dataset is  $256 \times 256 \times 3$  and  $256 \times 256 \times 1$ , respectively. The architecture consists of a total of 4 encoder blocks, 1 bottleneck block, and 4 decoder blocks. Each encoding block contains 2 Convolutional Layers, 2 Batch Normalization Layers with ReLU activation, and 1 MaxPool2D Layer. The bottleneck block has 1024 filters with a size of  $16 \times 16$ . The decoder block takes an input size of  $16 \times 16 \times 1024$ , increases the height and width of the image, and reduces the filter size as it performs upsampling. Each upsampling layer is concatenated with the corresponding Convolution block from the encoder.

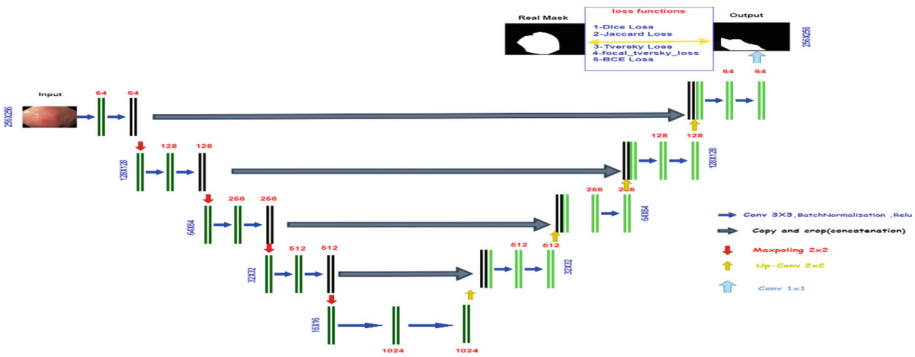


Fig. 2. Architecture of our Unet

## 5 Results and Discussions

Tables 2, 3, 4 and 5 summarize the results of the different models in the four test datasets of Kvasir-SEG, CVC-ClinicDB, ETIS-LaribPolypDB and the Combined data. By using roughly 200 annotated segmented examples, we achieved an optimal score Dice of 84.89% using Dice Loss (which corresponds to optimal IoU score of 76.73%). Focal Tversky loss and Jaccard Loss came in the second rank (Fig. 3).

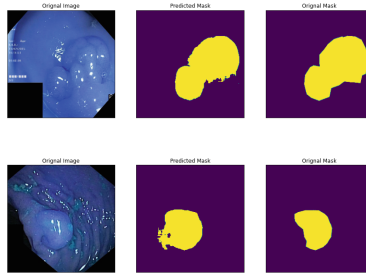
After training and testing on CVC-ClinicDB, we achieve the Result as shown in Table 3. By using roughly 123 annotated segmented examples, we achieved an optimal score Dice of 85.89% using Dice loss. Jaccard Loss comes in the second rank (Fig. 4).

By using roughly 40 annotated segmented examples and after training and testing on ETIS-LaribPolypDB I, we achieved an optimal score Dice of 77.02% using Tversky Loss. The difference from the previous scores obtained with the two last datasets and this one can be caused by the size of polyps as shown in Table 1 (Fig. 5).

After training and testing on the last combined kvasir, ETIS-LaribPolypDB and CVC-ClinicDB dataset, we achieve the result as shown in Table 5. By using roughly

**Table 2.** U-Net. Results of the model trained with the different loss functions, reported over the test set of Kvasir. Best values are indicated in bold.

Network	Dice Score (%)	IoU (%)
U-Net with BCE loss	82.36	74.38
<b>U-Net with Dice loss</b>	<b>84.89</b>	<b>76.73</b>
U-Net with Jaccard loss	82.83	74.56
U-Net with Tversky loss	82.17	73.36
U-Net with Focal Tversky loss	82.96	74.80

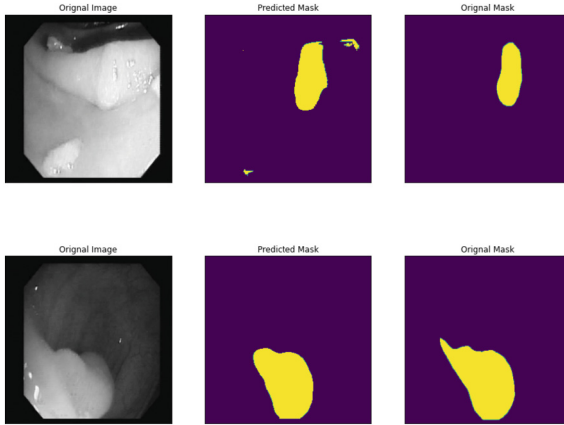


**Fig. 3.** Examples of segmentations in the test set of Kvasir produced by the U-net with Dice loss.

**Table 3.** U-Net. Results of the model trained with the different loss functions, reported over the test set of CVC-ClinicDB. Best values are indicated in bold.

Network	Dice Score (%)	IoU (%)
U-Net with BCE loss	73.72	64.73
<b>U-Net with Dice loss</b>	<b>85.89</b>	<b>79.13</b>
U-Net with Jaccard loss	84.03	76.33
U-Net with Tversky loss	81.75	73.36
U-Net with Focal Tversky loss	82.26	74.94

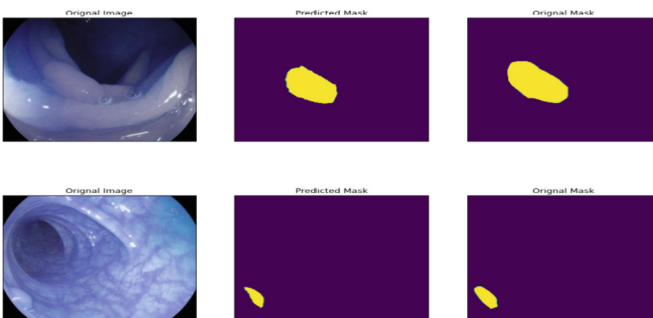
363 annotated segmented examples, we achieved an optimal score Dice of 76.71% using BCE Loss (Fig. 6).



**Fig. 4.** Examples of segmentations in the test set of CVC-ClinicDB produced by the U-net with BCE Loss.

**Table 4.** U-Net. Results of the model trained with the different loss functions, reported over the test set of ETIS-LaribPolypDB. Best values are indicated in bold.

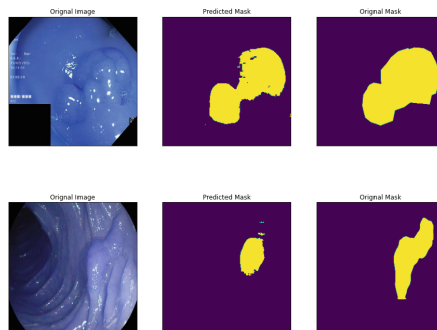
Network	Dice Score (%)	IoU (%)
U-Net with BCE loss	75.75	67.91
U-Net with Dice loss	75.26	68.00
U-Net with Jaccard loss	72.50	63.81
<b>U-Net with Tversky loss</b>	<b>77.02</b>	<b>70.10</b>
U-Net with Focal Tversky loss	62.43	55.60



**Fig. 5.** Example of segmentations in the test set of ETIS-LaribPolypDB produced by the U-net with Tversky Loss.

**Table 5.** U-Net. Results of the model trained with the different loss functions, reported over the test set of combined Data. Best values are indicated in bold.

Network	Dice Score (%)	IoU (%)
<b>U-Net with BCE loss</b>	<b>76.71</b>	<b>67.50</b>
U-Net with Dice loss	74.20	64.89
U-Net with Jaccard loss	74.03	65.33
U-Net with Tversky loss	68.31	58.98
U-Net with Focal Tversky loss	73.62	63.15



**Fig. 6.** Example of segmentations in the test set of combined data produced by the U-net with BCE Loss.

## 6 Conclusion

In this paper, we present a full procedure of images segmentation based on U-net architecture to label polyps. Experiments demonstrate that the loss function plays a crucial role in the model's performance. We use in this paper different metrics and loss function scores. Even if we obtained good results, based on the experimentation of different loss functions, this work is still under improvement. In fact, we can achieve better results by trying different architectures and including experiments with encoder features. In addition, we target in further research, to localize polyps, in real-time, from endoscopy videos frames. However, the videos processing is still a challenging task, comparatively to images, but results, if obtained from advanced models, will be more accurate.

## References

1. Selmouni, F., et al.: Delivering colorectal cancer screening integrated with primary health care services in Morocco: lessons learned from a demonstration project. *Cancer* **128**(6), 1219–1229 (2022)
2. Heresbach, D., et al.: Miss rate for colorectal neoplastic polyps: a prospective multicenter study of back-to-back video colonoscopies. *Endoscopy* **40**(04), 284–290 (2008). <https://doi.org/10.1055/s-2007-995618>



3. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. ArXiv (2015). Consulté le: 22 septembre 2022. <http://arxiv.org/abs/1505.04597>
4. Jha, D., et al.: Kvasir-SEG: a segmented polyp dataset. ArXiv (2019). Consulté le: 22 septembre 2022. <http://arxiv.org/abs/1911.07069>
5. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilarinho, F.: WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians. *Comput. Med. Imaging Graph.* **43**, 99–111 (2015). <https://doi.org/10.1016/j.compmedimag.2015.02.007>
6. Silva, J., Histace, A., Romain, O., Dray, X., Granado, B.: Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer. *Int. J CARS* **9**(2), 283–293 (2014). <https://doi.org/10.1007/s11548-013-0926-3>
7. Ruder, S.: An overview of gradient descent optimization algorithms. ArXiv (2017). Consulté le: 22 septembre 2022. <http://arxiv.org/abs/1609.04747>
8. Zhang, Z., Sabuncu, M.R.: Generalized cross entropy loss for training deep neural networks with noisy labels. ArXiv (2018). <https://doi.org/10.48550/arXiv.1805.07836>
9. Zou, K.H., et al.: Statistical validation of image segmentation quality based on a spatial overlap index. *Acad. Radiol.* **11**(2), 178–189 (2004). [https://doi.org/10.1016/S1076-6332\(03\)00671-8](https://doi.org/10.1016/S1076-6332(03)00671-8)
10. Jaccard, P.: The distribution of the flora in the alpine zone. *New Phytol.* **11**(2), 37–50 (1912). <https://doi.org/10.1111/j.1469-8137.1912.tb05611.x>
11. Tversky-features.pdf. Consulté le: 22 septembre 2022. <http://www.ai.mit.edu/projects/dm/Tversky-features.pdf>
12. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. ArXiv (2018). Consulté le: 22 septembre 2022. <http://arxiv.org/abs/1708.02002>
13. Shamir, R.R., Duchin, Y., Kim, J., Sapiro, G., Harel, N.: Continuous dice coefficient: a method for evaluating probabilistic segmentations. *Bioengineering* (2018). <https://doi.org/10.1101/306977>
14. Borgli, H., et al.: HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Sci. Data* **7**(1), 283 (2020). <https://doi.org/10.1038/s41597-020-00622-y>
15. Fernández-Esparrach, G., et al.: Exploring the clinical potential of an automatic colonic polyp detection method based on the creation of energy maps. *Endoscopy* **48**(09), 837–842 (2016). <https://doi.org/10.1055/s-0042-108434>
16. Lou, A., Guan, S., Ko, H., Loew, M.H.: CaraNet: context axial reverse attention network for segmentation of small medical objects. In: *Medical Imaging 2022: Image Processing*, San Diego, United States, p. 11 (2022). <https://doi.org/10.1117/12.2611802>



# Combination of an Improved Feistel Scheme and Genetic Operators for Chaotic Image Encryption

Hicham Rrghout<sup>1</sup> (✉), Mourad Kattass<sup>1</sup>, Mariem Jarjar<sup>1</sup>, Naima Benazzi<sup>2</sup>,  
Younes Qobbi<sup>1</sup>, Abdellatif Jarjar<sup>1</sup>, and Abdelhamid Benazzi<sup>1</sup>

<sup>1</sup> MATSI Laboratory, Mohammed First University, Oujda, Morocco  
{h.rrghout, mourad.kattass, qobbi.younes, a.benazzi}@ump.ac.ma

<sup>2</sup> LEEM Laboratory, Mohammed First University, Oujda, Morocco

**Abstract.** In this work, we will propose a new color image encryption technique using chaotic maps and based on an improved Feistel scheme. This scheme is reinforced by two genetic operators adapted to cryptography. After putting the original image into a single vector, a genetic mutation is applied. Then, a subdivision into sub-blocks of pseudo-random size is performed. Feistel's technique is applied using new confusion and diffusion functions. A genetic crossover is installed at the output and is controlled by a pseudo-random crossover table. Simulations carried out on a large sample of images of various sizes demonstrate that our method is resilient against all known attacks.

**Keywords:** Chaotic map · Feistel-round diagram · genetic operator

## 1 Introduction

With the immense development of the use of the Internet and its applications, data security becomes a real concern. The use of an encryption system makes it possible to remedy this problem, but the evolution of computing power threatens the confidentiality of these classic cryptographic systems. To overcome this concern, chaotic systems [1–4], are very efficient in this sector, thanks to the number of characteristics which present, in particular, sensitivity to initial conditions, and unpredictability. Due to the rapid growth of chaos theory in the field of mathematics, researchers now have the opportunity to greatly improve several established encryption techniques. In response to the heightened emphasis on security, numerous color image encryption schemes have emerged in the digital realm, most of which leverages chaos and number theory, notably Feistel [5–9]. Due to the confidentiality of digital medical images, researchers have developed algorithms that combine chaos and genetic operators [10–12] for powerful encryption systems.

### 1.1 Feistel’s Classical Technique

After separating the plain text into blocks of size  $2p$  with  $p \in \mathbb{N}$ , Feistel’s classical method [13] applied in the block  $(U_k = (G_k D_k))$  resumed by the scheme of the figure below (Fig. 1):

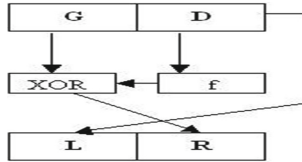


Fig. 1. Feistel’s classic round scheme

This encryption can be explained by the following expression, which works on two blocks  $(G_k, D_k)$ .

$$\left\{ \begin{array}{l} g_k(G_k, D_k) = (G'_k, D'_k) \\ \left\{ \begin{array}{l} G'_k = D_k \\ D'_k = G_k \oplus f_k(D_k) \end{array} \right. \end{array} \right. \text{ With } k \in \llbracket 1 \ t \rrbracket \quad (1)$$

$(f_k)$  is a pre-defined pseudo-random function. In the absence of any diffusion, these systems remain exposed to differential attacks. Several techniques to improve Feistel’s solution have invaded the network world Most of them are focused on increasing the number of turns.

### 1.2 Genetic Operator

Every genetic algorithm is built on a set of specific functions. The most popular and widely employed in cryptography are:

- acting on DNA:
  - o Relocation; Reversion; Mutation
- acting on Codon RNA
  - o Crossover

In our approach, we will apply a mutation at re-entry and a crossover at the exit of Feistel’s lap.

## 2 Related Works

Currently, chaotic cryptography is widely used to secure the transmission of multimedia data over the internet, and the combination between chaos and Feistel’s technique and genetic operators represents a new technique for image encryption. Several studies

have been based on this approach. The authors of the article [14] described a new technique that begins with the application of an advanced Feistel scheme and ends with the implementation of deeply modified genetic operators. After vectorization of the original image, an application of the advanced Feistel scheme on blocks of random size will be launched. The output vector is transcribed into restricted ASCII code to perform genetic cross-matching suitable for color image encryption. The resulting output vector is transcribed in restricted ASCII code to greatly enhance the effects of genetic crossover. The results show the guarantee of this approach against all known attacks. Man Kang et al. [15] proposed a new genetic algorithm. The new genetic algorithm generates  $8 \times 8$  bijective S-boxes with 6 differential uniformity, 108 nonlinearity and 10 boomerang uniformity, which significantly improved the properties of the S-boxes created by the Feistel structure. Moreover, the new genetic algorithm also improves the properties of the S-box population created by the Feistel structure as a whole. The results show that the S-boxes generated by the new genetic algorithm have better properties than the S-boxes generated by the traditional genetic algorithm, demonstrating the efficiency and superiority of the new genetic algorithm in S-box development. The authors of [16] studied an encryption scheme which mainly adopts four encryption steps to encrypt the simple image, which are Chaotic Sequence Generation, Hill Encryption, Feistel Network and Pixel Broadcast. Analyzes show that there are some problems in the design of the secret key and the encryption process of this encryption scheme. After reporting and analyzing these issues, they made several necessary improvements to this encryption scheme and proposed the corresponding chosen plaintext attack algorithm. Simulation tests and subsequent analyzes confirmed the effectiveness and feasibility of the proposed attack algorithm.

### 3 The Proposed Method Description

Our novel approach, which is based on the application of chaos and confused diffusion, significantly improves Feistel's scheme with the addition of two genetic operators that are particularly well-suited to color image encryption. The following stages serve as the foundation for this novel method.

#### 3.1 Chaotic Sequences Development

The choice of the two chaotic maps for the development of the keys and subkeys of our algorithm is related to the ease of their handling and to the high sensitivity to the initial conditions.

##### 3.1.1 The Chaotic Map (PWLCM)

The chaotic map (PWLCM) [17] is composed of several linear segments defined by Eq. (2):

$$u_n = F(u_{n-1}, d) = \begin{cases} \frac{u_{n-1}}{d}, & 0 \leq u_{n-1} \leq d \\ \frac{u_{n-1}-d}{0.5-d}, & d \leq u_{n-1} \leq 0.5 \\ F(1 - u_{n-1}, d), & d \leq u_{n-1} \leq 1 \\ u_0 \in ]0; 1[ \text{ and } d \in ]0; 0.5[ \end{cases} \quad (2)$$

### 3.1.2 Skew Tent Map (SKTM)

The following equation [18] presents the Skew Tent map ( $v_n$ ) expression:

$$\left\{ \begin{array}{l} v_0 \in ]0, 1[ \quad p \in ]0, 5[ \\ v_{n+1} = \begin{cases} \frac{v_n}{p} & \text{if } 0 < v_n < p \\ \frac{1-v_n}{1-p} & \text{if } p < v_n < 1 \end{cases} \end{array} \right. \quad (3)$$

The combination of these two chaotic maps makes it possible to generate pseudo-random vectors that we will use in our approach.

### 3.1.3 Generation of Chaotic Vectors

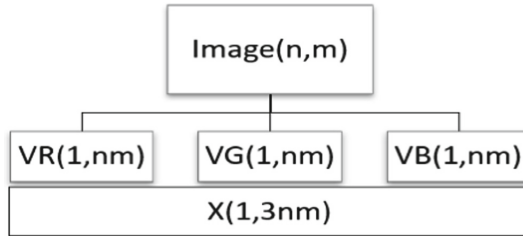
Our work requires the construction of five chaotic confusion vectors (C1), (C2), (C3), (C4) and (C5) with a coefficient of (G\_256) and two binary control vectors (B1) and (B2).

Algorithm1: Chaotic used vector design

1. For  $i = 1$  to  $3nm$
2.  $C1(i) = \text{mod}(E(|u(i) - v(i)| * 10^{10}, 253) + 3)$
3.  $C2(i) = \text{mod}(E(u(i) * v(i)) * 10^{10}, 253) + 2)$
4.  $C3(i) = \text{mod}(E(u(i) * 10^{11}, 253) + 1)$
5.  $C4(i) = \text{mod}(E(v(i) * 10^{10}, 253) + 1)$
6.  $C5(i) = \text{sup}(C3(i); C5(i))$
7.  $TC(i) = \text{Mod}(C3(i); 3m)$
8.  $TT(i) = \text{Mod}(C4(i); 3m)$
9.  $DV(i) = \text{Mod}(C4(i); 3) + 1$
10. If  $u(i) \geq v(i)$  then  $B1(i) = 0$  else  $B1(i) = 1$  end if
11. If  $C2(i) > C4(i)$  Then  $B2(i) = 0$  else  $B2(i) = 1$  end if
12. Next  $i$

### 3.2 Vectorization of the Original Image

The vectorization of an image of size (n,m) goes through the extraction of the three channels (R, G, B), then a concatenation in the form of a vector X of size 3 nm.



This operation can be described using the following algorithm:

---

Algorithm2: Vectorization of the original image

---

1. **For** i = 0 to nm-1
  2. **If** B1(i) = 0 **Then**
  3.  $X(3i) = Vr(i) \oplus C1(i)$ ;  $X(3i + 1) = Vg(i) \oplus C2(i)$ ;  $X(3i + 2) = Vb(i) \oplus C3(i)$
  4. **Else**
  5.  $X(3i) = Vr(i) \oplus C4(i)$  ;  $X(3i + 1) = Vg(i) \oplus C1(i)$  ;  $X(3i + 2) = Vb(i) \oplus C5(i)$
  - Next** i
- 

The final image has a flat histogram because this transition excludes any correlation between pixel intensity.

### 3.3 Genetic Mutation Application

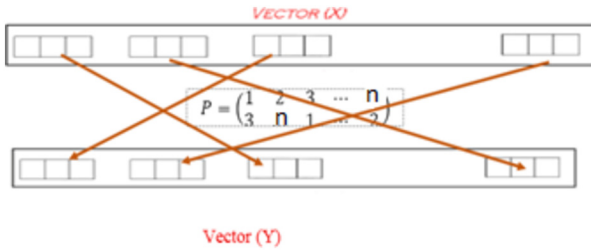
The vector X is subdivided into n sub-vectors, each of size 3 m (Fig. 2).



**Fig. 2.** Block partitioning of (3 m) pixels

At the level of the blocs on the vector, we apply the permutation (PB) obtained by sorting the (n) input data of the chaotic map PWLCM (u).

**Example**



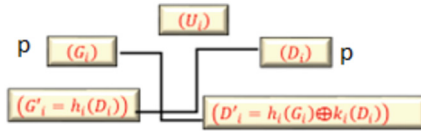
**Fig. 3.** Feistel’s diagram

The output vector (Y) is also subdivided into (n) sub-blocks of size (3m) each. The diagram in the image below offers as an illustration with this Feistel tower (Fig. 3).

**Block Size Adaptation**

To adapt the size of the block, use the following formulas:

Let  $r = 3m \% 2$ ;  $t = 3m - r$  and  $p = t / 2$  Then:  $3m = 2p + r$  With  $U_i$  is a block of index  $i$  (Fig. 4).



**Fig.4.** New Feistel Scheme

$(h_i)$  is a chaotic displacement of the pixels of the block, of step TC(i) and  $(k_i)$  is a chaotic displacement of the pixels of the block, of step TT(i). This scheme is interpreted by the algorithm below:

---

Algorithm3: Feistel circuit

---

1. **For**  $i = 1$  **to**  $p - 1$
2.  $G'_k(i) = h_i(D_i)$
3.  $D'_k(i) = h_i(G_i) \oplus k_i(D_i)$
4. **if**  $r=1$
5.  $U_i(3m-1) = U_i(3m-1) \oplus C1(n+i)$
6. **Next**  $i$

---

### Diffusion Phase

Diffusion is a cryptographic method developed with the purpose of enhancing the redundancy of plain text, thereby concealing the statistical patterns inherent in the text.

- **Initialization vector S:**

To determine the initialization vector, the following algorithm is used:

---

Algorithm4: Initialization vector

---

```

For j=0 to n-1
  For i=0 to 3m-1
    M(j,i)=X(3mj+i)
  For i=1 to n-1
    For j=0 to 3m-1
      S(j)= S(j) ⊕ M(i,j)

```

---

- **Modification of the first block**

```

For i = 0 to 3m-1
  U(i) = U(i) ⊕ S(i)

```

### Genetic Cross-Over

The output vector will be subjected to a genetic cross with the confusion vectors (C1)(C2)and(C3) under the control of the vector (DV). The Crossover function is described by the following algorithm:

---

Algorithm 5: Genetic Crosse-over

---

1. for i=0 to 3nm-1
2. If DV(i)=1 Then
3. Z(i)=Y(i)⊕C1(i)
4. If DV(i)=2 Then
5. Z(i)=Y(i)⊕C2(i)
6. Else
7. Z(i)=Y(i)⊕C3(i)
8. Next i

---

The encrypted image is represented by the vector Z.

## 4 Examples and Simulations

In this section, an extensive number of medical images and reference colors are chosen to be tested later by our algorithm. First, a simulation study is described for testing a few images.



### 4.1 Brutal Assaults

The process involves randomly reconstructing the encryption keys.

#### Key-Space Analysis

The key space of a cryptosystem plays a vital role in assessing its vulnerability to brute force attacks. For a robust algorithm, the key space should be maximized and exceed  $2^{100}$ . In our suggested approach, we employed four parameters, each encoded with 32 bits. Consequently, the overall key space for our key amounts to  $2^{128}$ . we deduce that our approach is secure against any brute force attack.






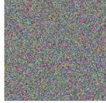
### 4.2 Statistics Attack Security

To demonstrate the robustness of our novel medical and color image encryption technique against statistical attacks, extensive testing has been conducted on various images. Here, we highlight the most notable tests performed:

#### Histogram Analysis

An image histogram is a graphical representation that displays the number of pixels with the same gray level in an image. From a cryptography standpoint, an encrypted image’s consistent color distribution is critical and provides good protection against statistical assaults. Table 1 shows some of our system’s simulation outcomes.

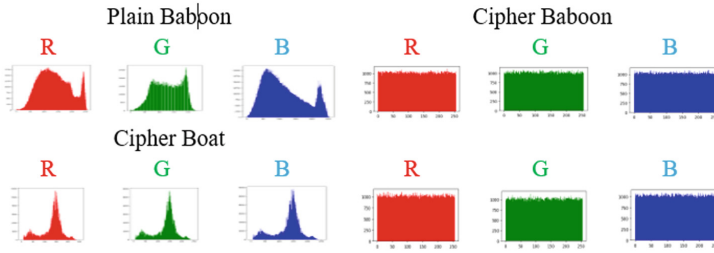
**Table 1.** Visual test analysis

Name	Plain image	Encrypted Image
Baboon 512 ×512		
Boat 512 ×512		
Cameramen 512 ×512		

The remarkable resilience to statistical attacks is demonstrated by the uniform distribution of the histograms of the numerical images. The diagram below exemplifies this (Table 2).

In this study, we observed that our algorithm consistently produces encrypted images from reference images with flattened histograms. This uniformity in the histogram distribution provides significant protection against histogram-based attacks.

**Table 2.** Histogram analysis



**Entropy Analysis**

Founded in 1949 by Claude E. Shannon. The modern theory of information is interested in the correction of errors, in the protection against attacks. The entropy of an image of size  $(n, m)$  [7] is given by this equation.

$$H(MC) = \frac{1}{t} \sum_{i=1}^t -p(i) \log_2(p(i)) \tag{4}$$

The probability of occurrence of level  $i$  in the original image is denoted by  $p(i)$ . Table 3 presents the entropy values of the images tested using our technique.

**Table 3.** Entropy analysis

Entropy Values									
Name	Original image			Cipher image			8 - Cipher Entropy		
	Red	Green	Blue	Red	Green	Blue	Red	Green	Blue
Baboon	7.998	7.999	7.998	7.9988	7.9992	7.9987	$\approx 0$	$\approx 0$	$\approx 0$
Cameramen	7.998	7.999	7.998	7.999	7.9991	7.9989	$\approx 0$	$\approx 0$	$\approx 0$

**Correlation Analysis**

Correlation is the search for a simple linear relationship between adjacent pixels [19].

The correlation of an image with dimensions  $(n,m)$  can be calculated using the following equation:

$$r = \frac{\text{cov}(x, y)}{\sqrt{V(x)}\sqrt{V(y)}} \tag{5}$$

The table below compares the original image’s pixel correlations to the encrypted image’s pixel correlations.

The Table below gives the values of the correlation coefficients in the three directions (vertical, horizontal and diagonal) (Table 4):

**Table 4.** Correlation analysis

Image	channel	direction	Correlation coefficients
Cameramen	Red	H	-0.0026
		V	0.0008
		D	0.0015
	Green	H	-0.0019
		V	-0.0017
		D	-0.0023
	Blue	H	-0.0026
		V	-0.0018
		D	-0.0002
Boat	Red	H	0.0021
		V	-0.0029
		D	-0.0016
	Green	H	0.0016
		V	-0.0029
		D	0.0043
	Blue	H	-0.0016
		V	-0.0009
		D	-0.0004

We can see that the correlation value is near to zero, which means that we are better protected from correlation assaults.

### 4.3 Differential Analysis

Consider two encrypted images, denoted as  $C_1$  and  $C_2$ , where their corresponding plaintext images differ by only one pixel [18]. The mathematical analysis of NPCR (Normalized Pixel Change Rate) for an image can be expressed by the following equation:

$$\text{NPCR} = \left( \frac{1}{nm} \sum_{i,j=1}^{nm} D(i, j) \right) * 100 \quad (6)$$





$$\text{With } D(i, j) = \begin{cases} & \&1 & \text{if } C_1(i, j) \neq C_2(i, j) \\ & \&0 & \text{if } C_1(i, j) = C_2(i, j) \end{cases} \quad (7)$$

The UACI mathematical analysis of an image can be expressed by the following equation:

$$\text{UACI} = \left( \frac{1}{nm} \sum_{i,j=1}^{nm} \text{Abs}(C_1(i, j) - C_2(i, j)) \right) * 100 \quad (8)$$

The differential constants found in our simulations performed on a set of reference images are shown in the table below (Table 5):

**Table 5.** Differential analysis

Image				
NPCR	99.62	99.61	99.60	99.63
UACI	33.46	33.42	33.44	33.45

#### 4.4 Comparison

The following table presents the results of the comparison of the performance of our method with several other techniques based on the values of entropy, NPCR, UACI, correlation coefficient (Table 6).

**Table 6.** Comparison with other approaches

	Image	Our approach	Réf [20]	Réf [21]
Entropy	Baboon $512 \times 512$	7.9998	7.9987	7.9997
NPCR	Baboon $512 \times 512$	99.62	99.57	99.63
UACI	Baboon $512 \times 512$	33.46	33.17	33.43
Vertical Correlation	Baboon $512 \times 512$	-0.0022	-0.0002	0.0061

## 5 Conclusion

Several cryptosystem approaches are proposed in the field of image encryption, but with the rapid development of technology, it is necessary to build more efficient cryptosystems to overcome some weaknesses. Our approach proposed in this article is based on a profound improvement of the Feistel scheme. This improvement consists in using two transformations controlled by chaotic vectors inside a Feistel round. Our approach is enhanced by two genetic operators, which presents a high level of complexity due to the mutation, which consists of a chaotic exchange of somatic blocks under the supervision of a decision vector and a diffusion phase, which allows to increase the complexity of our algorithm and also to protect it against differential attacks.






## References

1. Khan, M.F., Ahmed, A., Saleem, K., Shah, T.: A novel design of cryptographic sp-network based on gold sequences and chaotic logistic tent system. *IEEE Access* **7**, 84980–84991 (2019). <https://doi.org/10.1109/ACCESS.2019.2925081>
2. Tutueva, A.V., Nepomuceno, E.G., Karimov, A.I., Andreev, V.S., Butusov, D.N.: Adaptive chaotic maps and their application to pseudo-random numbers generation. *Chaos Solitons Fractals* **133**, 109615 (2020). <https://doi.org/10.1016/j.chaos.2020.109615>
3. Jamal, S.S., Anees, A., Ahmad, M., Khan, M.F., Hussain, I.: Construction of cryptographic S-boxes based on mobius transformation and chaotic tent-sine system. *IEEE Access* **7**, 173273–173285 (2019). <https://doi.org/10.1109/ACCESS.2019.2956385>
4. Teh, J.S., Alawida, M., Sii, Y.C.: Implementation and practical problems of chaos-based cryptography revisited. *J. Inf. Secur. Appl.* **50**, 102421 (2020). <https://doi.org/10.1016/j.jisa.2019.102421>
5. JarJar, A.: Improvement of Feistel method and the new encryption scheme. *Optik* **157**, 1319–1324 (2018)
6. Lallemand, V., Minier, M., Rouquette, L.: Automatic search of rectangle attacks on feistel ciphers: application to WARP. *IACR Trans. Symm. Cryptol.* 113-140 (2022). <https://doi.org/10.46586/tosc.v2022.i2.113-140>
7. Al-Bahrani, E.A., Kadhum, R.N.: A new cipher based on Feistel structure and chaotic maps. *Baghdad Sci. J.* **16**(1), 270–280 (2019). [https://doi.org/10.21123/bsj.2019.16.1\(Suppl.\).0270](https://doi.org/10.21123/bsj.2019.16.1(Suppl.).0270)
8. Abd Ali, S.M., Hasan, H.F.: Novel encryption algorithm for securing sensitive information based on feistel cipher. *Test Eng. Manag.* **19**(80), 10–16 (2019)
9. Ito, G., Iwata, T.: Quantum distinguishing attacks against type-1 generalized feistel ciphers (2019). Consulté le: 25 avril 2023. <https://eprint.iacr.org/2019/327>
10. Qobbi, Y., Jarjar, A., Essaid, M., Benazzi, A.: Image encryption algorithm based on genetic operations and chaotic DNA encoding. *Soft Comput.* **26**(12), 5823–5832 (2022). <https://doi.org/10.1007/s00500-021-06567-7>
11. Lambora, A., Gupta, K., Chopra, K.: Genetic algorithm- a literature review. In: 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), pp. 380-384 (2019). <https://doi.org/10.1109/COMITCon.2019.8862255>
12. Zhou, Y., et al.: Hybrid genetic algorithm method for efficient and robust evaluation of remaining useful life of supercapacitors. *Appl. Energy* **260**, 114169 (2020). <https://doi.org/10.1016/j.apenergy.2019.114169>
13. Ahmad, M., et al.: An image encryption algorithm based on new generalized fusion fractal structure. *Inf. Sci.* **592**, 1–20 (2022). <https://doi.org/10.1016/j.ins.2022.01.042>
14. Hraoui, S., JarJar, A.: Single Feistel lapse acting on reduced ASCII codes followed by a genetic crossover. *SN Appl. Sci.* **4**(4), 1–18 (2022)
15. Kang, M., Wang, M.: New genetic operators for developing S-boxes with low boomerang uniformity. *IEEE Access* **10**, 10898–10906 (2022). <https://doi.org/10.1109/ACCESS.2022.3144458>
16. Feng, W., Qin, Z., Zhang, J., Ahmad, M.: Cryptanalysis and improvement of the image encryption scheme based on feistel network and dynamic DNA encoding. *IEEE Access* **9**, 145459–145470 (2021). <https://doi.org/10.1109/ACCESS.2021.3123571>
17. Manihira, N.R., Dauda, A.K.: Image encryption using chaotic maps and DNA encoding. *Int. J. Eng. Res. Technol.* **10**(11), 1–5 (2022). <https://doi.org/10.17577/IJERTCONV10IS11137>
18. Ghoul, S., Sulaiman, R., Shukur, Z.: A review on security techniques in image steganography. *Int. J. Adv. Comput. Sci. Appl.* **14**(6), 1–25 (2023). <https://doi.org/10.14569/IJACSA.2023.0140640>

19. Zhang, M., Li, Y., Song, H., Wang, B., Zhao, Y., Zhang, J.: Security analysis of quantum noise stream cipher under fast correlation attack. In: Optical Fiber Communication Conference (OFC) 2021 (2021), paper Th1A.5, p. Th1A.5. Optica Publishing Group (2021). <https://doi.org/10.1364/OFC.2021.Th1A.5>
20. Ghazvini, M., Mirzadi, M., Parvar, N.: A modified method for image encryption based on chaotic map and genetic algorithm. *Multimedia Tools Appl.* **79**, 26927–26950 (2020). <https://doi.org/10.1007/s11042-020-09058-3>
21. Essaid, M., Akharraz, I., Saaidi, A.: Image encryption scheme based on a new secure variant of Hill cipher and 1D chaotic maps. *J. Inf. Secur. Appl.* **47**, 173–187 (2019)



# Post-processing of Closed Contours to Obtain Inscribed $K$ -Sided Polygons

R. Molano<sup>1</sup>(✉) , M. Ávila<sup>2</sup> , J. C. Sancho<sup>2</sup> , P. G. Rodríguez<sup>2</sup> ,  
and A. Caro<sup>2</sup> 

<sup>1</sup> Department of Mathematics, Universidad de Extremadura, 10003 Cáceres, Spain  
rmolano@unex.es

<sup>2</sup> Department of Computer and Telematics Systems Engineering,  
Universidad de Extremadura, 10003 Cáceres, Spain  
{mmavila, jcsanchon, pablogr, andresc}@unex.es

**Abstract.** This paper presents an algorithm for determining the maximum area or perimeter of a simple or convex polygon with  $k$  sides located within a region of interest. Inscribed polygons are used to detect, classify, and segment objects in images. Traditional algorithms generate basic shapes such as triangles, quadrilaterals, and pentagons. However, a user may be interested in obtaining polygons with the maximum area or perimeter, deciding whether they must be simple or convex, and even specifying the type of polygon needed (triangle, quadrilateral, etc.). The problem of determining the maximum area or perimeter of a simple  $k$ -sided polygon within a region of interest has not been solved in a generic way for a user-defined value of  $k$ . This paper presents a flexible algorithm that allows the user to determine which polygons should be calculated. The algorithm's source code in C++, Java and Python is available in a GitHub repository to ensure its usability for the scientific community.

**Keywords:** Region of Interest ·  $k$ -sided · Polygon · Area · Perimeter

## 1 Introduction

Obtaining inscribed polygons within a closed contour can be useful in multiple areas, such as geometry, physics, engineering, computer science, among others. In geometry, the inscribed polygon can be used to calculate the area of the polygon and has many applications in surveying, urban planning, and construction. In physics, it can be used to analyze object motion trajectories and calculate velocities, accelerations, energies, among other variables. In engineering, it can be used for resource optimization, such as cutting a sheet of material into several pieces, minimizing waste. In object design, the inscribed polygon can represent the cross-sectional area of a mechanical part within the available space.

In image analysis and pattern recognition, the inscribed polygon can be used to detect and classify objects in an image, measure their size and shape, segment objects from the background, and reduce noise in images. The inscribed polygon

can also be used for feature detection, such as vertex angles and object symmetry, allowing object identification and classification.

Practical applications of obtaining inscribed polygons include crop and pasture analysis, land management planning, monitoring crop growth and yield, and animal tracking in agriculture and livestock. In medicine, the inscribed polygon can be used for medical image analysis, identifying regions of interest or validating image processing algorithms. In the photovoltaic solar energy industry, analyzing regions of interest in satellite images can optimize the placement of solar panels on rural farms and on the roofs of houses and buildings in cities. For urban planning of a city, it can be used to select the best region to build a hotel, a residential building, or a children's playground.

Polygons are the most effective way to approximate irregular regions of interest (ROI) or closed contours. Traditional solutions, however, provide limited options with basic shapes such as triangles, quadrilaterals, pentagons, etc., and they are unable to incorporate new constraints. It should be possible for users to choose whether they wish to calculate a simple or convex polygon, as well as what type of polygon they wish to calculate (triangle, quadrilateral, pentagon, etc.).

A solution has not yet been proposed to this problem, since no article has yet been published that computes the simple or convex  $k$ -sided polygon with the maximum area or perimeter within a ROI given a user-defined value  $k$ . With the method proposed in this paper, any researcher can obtain the polygon that best fits their needs. Instead of using different algorithms to compute the rectangle with the maximum-area or the hexagon with the maximum-perimeter, this approach provides a comprehensive solution that offers all possible options in an easy-to-use manner.

Computational geometry has been used to solve problems utilizing geometric optimization techniques in order to obtain all these polygons. Thus, the goal is to determine the optimal solution within a set of all possible solutions to a given problem. In terms of area and for triangles, Der Hoog et al. [4] demonstrated how to compute the maximum-area triangle in a convex polygon can be computed in  $O(n \log n)$  time. Later, Kallus [6] solved this problem in linear time ( $O(n)$ ), and Melissarattos and Souvaine [10] presented an algorithm for finding the maximum-area triangle inscribed in a simple  $n$ -sided polygon in  $O(n^4)$  time.

For rectangles, Alt, Hsu and Snoeyink [2] computed the largest area axis-parallel rectangle in a convex polygon in  $O(\log n)$  time and Boland and Urrutia [3] solved the problem in  $O(n \log n)$  time for a simple polygon. Removing the axis-aligned condition, Kanuer et al. [8] considered approximation algorithms and proved that the rectangle with the largest area in a convex polygon could be computed in  $O(\frac{1}{\epsilon} \log \frac{1}{\epsilon} \log n)$  time. For simple polygons, Molano et al. [12] solved the problem in  $O(n^3)$  time, showing how to compute the largest area rectangle of arbitrary orientation in a closed contour working with ROIs. For parallelograms, Jin [5] found an algorithm to compute the largest parallelogram inside a convex polygon in  $O(n \log^2 n)$  time and Molano et al. [11] solved the same problem in  $O(n^3)$  time but for simple polygons. Finally, Keikha [7] and Rote [13] developed an algorithm to compute the quadrilateral of largest area contained in a convex polygon, in  $O(n)$  time.



Based on the literature, two conclusions can be drawn. Firstly, there is a significant difference in computational cost when the initial polygon is changed from convex to nonconvex. Secondly, there is no generic solution that can calculate any  $k$ -sided polygon and, therefore, be adapted to meet the needs of every user. In this paper, the goal is to develop an algorithm that computes a convex or simple  $k$ -sided polygon of maximum area or perimeter within any closed contour given a fixed number  $k$ . To ensure the usability of our method, we show the algorithm in pseudocode and define it into two subprograms that can easily be incorporated into future research projects by any interested researcher.

The following sections provide a detailed description of the algorithm described in this paper. The algorithm is presented in Sect. 2. In Sect. 3, we present a variety of results related to the practical application of our proposal. The conclusion of our research is presented in Sect. 4.

## 2 Methodology

The Algorithm 2 *Solution( $N$ , distance, matrix, function)* is the main program of the proposal. It is composed of four parameters:  $N = \#P$ , number of points of the lattice polygon  $P$ ;  $k = \text{distance} + 1$ , where  $k$  is the number of sides of the polygon to be computed; matrix, which is the adjacency matrix; and function, which indicates whether to calculate the area (function = 0) or the perimeter (function = 1). In addition, by making some modifications to Algorithm 2 it is possible to calculate all the solutions for convex polygons, *Solution-conv( $N$ , distance, matrix, function)*.

To obtain the solutions by means of Algorithm 2, we follow these five steps:

1. Select the ROI and specify the closed contour  $C$  in which it is included.
2. Construct the lattice polygon  $P$  within the closed contour  $C$ . We define a *lattice polygon* as a polygon whose points have integer coordinates and which has the following properties: it is defined on a *regular partition*  $\Pi$ , the connections between consecutive vertices are not necessarily established in the eight directions ( $\pi k/4, k = 0, \dots, 7$ ), and the edges of the polygon do not intersect except at their vertices. In addition, we define *partition size*,  $L = |x_{i+1} - x_i| = |y_{j+1} - y_j|$ , the length of the side of each square formed by the square grid and we state that partition  $\dot{\Pi}$  is finer than partition  $\Pi$ , if it is verified that all points of  $\Pi$  belong to  $\dot{\Pi}$ . We denote  $\Pi \preceq \dot{\Pi}$ .

If  $\#$  represents the cardinality of the set and  $P = \partial P \cup \iota P$ , where  $\partial P$  is the family consisting of boundary nodes of  $P$ , and its complementary in  $P$ ,  $\iota P$ , the interior points, then  $\#(P) = N = n + o \simeq \lambda n, \lambda \in \mathbb{N}$  with  $\#(\partial P) = n$  and  $\#(\iota P) = o$ .

3. Construct the *adjacency matrix*  $A = (a_{ij})$  given the  $N$  points of the lattice polygon  $P$ , where  $a_{ij} = 1$  if there is an edge between  $i$  and  $j$  and 0, otherwise.
4. Compute the polygons that are a certain distance from two points, point1 and point2 (Algorithm 1), given a partition size  $L$ .

Algorithm 1 starts by considering that the two points are connected (Line 2). Then, it calculates all edges that are at distance 1 from point1 and stores

them successively until the chosen distance is reached (Lines 3–4). The computational cost is determined by the loop (Line 3) in  $O(k)$  time and Line 4 in  $O(n^3)$  time [9]. Thus, the computational cost of the loop of Lines 3–4 is  $O(n^3k)$ . Finally, the algorithm eliminates any edges that cannot form a polygon through two auxiliary functions, *Segments* and *Aligned* (Lines 6–8), with computational cost  $O(k^2)$  and  $O(k)$ , respectively [9]. The first one determines whether two segments intersect or not, and the second one, if three consecutive vertices for a proposed solution are aligned or not. The time complexity of this process is determined by the loop (Line 6), with a time complexity of  $O(n^2)$ , and the above mentioned functions. Therefore, the computational cost of Lines 6–8 is  $O(n^2k^3)$ . Consequently, the computational cost of Algorithm 1 is directly related to the maximum cost of the two loops discussed above. In other words,  $Max(n^3k, n^2k^3) = n^3k$  if we assume that  $n \gg k$ .

---

**Algorithm 1:** *Polygons(point1, point2, distance, matrix)*

---

```

1 edges ← {point1}
2 if matrix(point1, point2) = 1 then
3   foreach i from 1 to distance
4     Compute the edges of all possible polygons from point1
5   Initialize sides
6   foreach i from 1 to Length(edges)
7     if edges[i][distance + 1] = point2 and Segments(edges[i]) = False and
      Aligned(edges[i]) ≠ 0 then
8       Insert edges[i] in sides
9 return sides

```

---

5. Compute the simple  $k$ -sided polygon of a maximum-area or perimeter inscribed in a lattice polygon (Algorithm 2).

Algorithm 2 starts by calculating all the simple  $k$ -sided polygons (Lines 2–5), and then the functions *Update* and *Duplicate* are applied (Line 6). The first function, computes the largest area or perimeter polygon depending on the value of the "function" parameter (0 for the largest area and 1 for the perimeter), and the second function eliminates repeated solutions. The computational cost of the above functions are  $O(n^2k)$  and  $O(n^4k \log k)$ , respectively [9]. The computational cost of Algorithm 2 is determined by two nested loops, which run in  $O(n^2)$  time, and Algorithm 1 (Line 4), which runs in  $O(n^3k)$ . Therefore, the time complexity is  $O(n^5k)$ , because  $Max(n^5k, n^2k, n^4k \log k) = n^5k$ .

The source code developed in C++, Java and Python is available in a GitHub repository for research purposes [9]. Thus, any researcher is able to thoroughly examine the code, understand the inner workings of the algorithms and functions, and potentially make modifications or improvements to the code.

---

**Algorithm 2:** *Solution*( $N$ , *distance*, *matrix*, *function*)

---

```

1 Initialize solution, sides
2 foreach i from 1 to N-1
3   foreach j from i+1 to N
4     if Polygons (i, j, distance, matrix)  $\neq 0$  then // (Alg. 1)
5       Insert Polygons (i, j, distance, matrix) in sides
6 solution  $\leftarrow$  Duplicates (Update (sides, function))
7 return solution

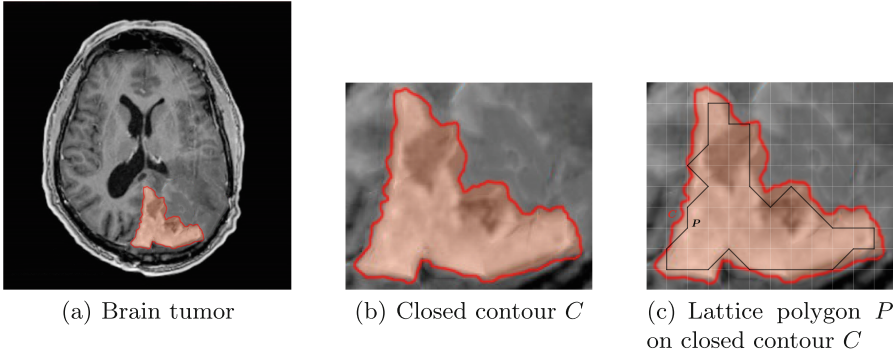
```

---

### 3 Results and Discussion

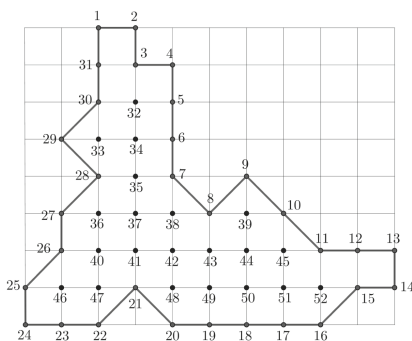
#### 3.1 Medical Imaging

From the National Cancer Institute (NIH), the image of a brain tumor has been obtained, specifically, a meningioma (Fig. 1a), which is a primary tumor of the central nervous system (CNS) that originates in the brain or spinal cord. Once these images are obtained, the closed contour  $C$  is created (Fig. 1b) generating a collection of points. Finally, the lattice polygon  $P$  (Fig. 1c) necessary to apply the Algorithm 2 is built, where  $N = \#(P) = 52$  and  $P = \{p_1, \dots, p_{52}\}$ .



**Fig. 1.** Example 1: image extraction.

We have included Table 1 that display some of the solutions obtained from Fig. 1c (Fig. 2) using Algorithm 2. Additionally, we have included Table 2 to calculate solutions for convex  $k$ -sided polygons. By comparing the solutions for area and perimeter obtained from nonconvex and convex polygons, we observe the differences between the two types of polygons.



**Fig. 2.** Example 1: lattice polygon.

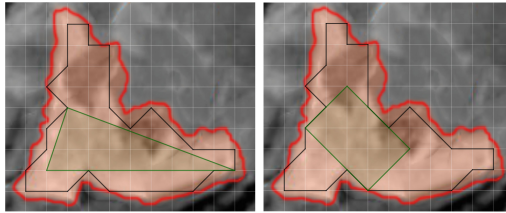
**Table 1.** Example 1: maximum-area and perimeter simple  $k$ -sided polygon.

distance	polygon	<i>Solution(52, distance, matrix, 0)</i>
2	Triangle	(14,28,46), (15,35,25)
3	Quadrilateral	(3,46,14,38), (16,30,40,20), (16,34,27,20), (16,34,24,21)
	Rectangle	(4,31,47,48), (27,20,44,34), (10,27,46,51)
4	Pentagon	(1,47,15,9,43)
5	Hexagon	(1,43,9,15,20,40), (1,43,13,16,20,40), (4,42,9,15,19,26), (13,43,32,24,21,16)
6	Heptagon	(1,40,20,15,9,42,5), (1,43,13,16,21,24,28), (1,43,9,15,16,20,40), (4,42,9,15,16,21,24), (4,42,9,15,17,21,24), (4,42,9,15,18,21,24), (4,42,9,15,19,21,24), (4,42,9,15,20,21,24), (4,42,9,15,20,40,31), (9,43,32,24,21,16,15), (9,43,32,24,21,17,15), (9,43,32,24,21,18,15), (9,43,32,24,21,19,15), (9,43,32,24,21,20,15), (13,43,32,24,21,20,16)
distance	polygon	<i>Solution(52, distance, matrix, 1)</i>
2	Triangle	(14,36,25)
3	Quadrilateral	(14,28,45,25)
4	Pentagon	(14,36,15,46,28)
5	Hexagon	(4,46,15,40,14,41)
6	Heptagon	(4,46,12,47,13,21,24)

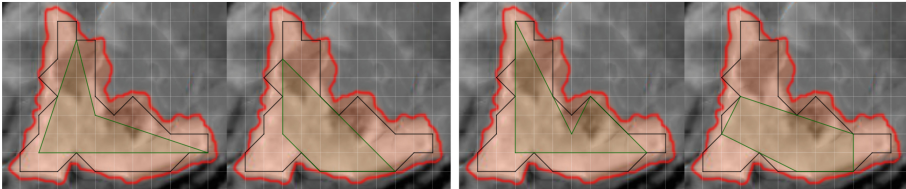
The solutions presented in Fig. 3 were generated using Algorithm 2. As the number of sides increases, the area also correspondingly increases. Furthermore, to demonstrate the flexibility of the proposed algorithm, we have included the solution for obtaining the rectangle with maximum area, along with several convex polygons that can be attained by making slight modifications to the main algorithm.

**Table 2.** Example 1: maximum-area and perimeter convex  $k$ -sided polygon.

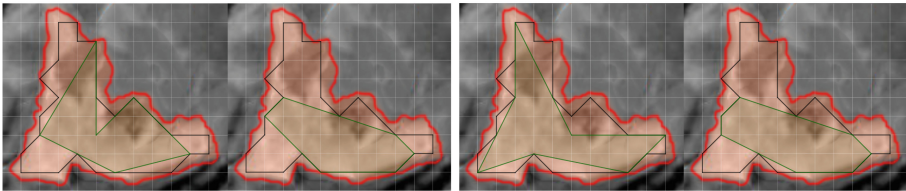
distance	polygon	$Solution-conv(52, distance, matrix, 0)$
3	Quadrilateral	(16,30,40,20), (16,34,27,20)
4	Pentagon	(10,27,20,16,15), (11,28,26,19,16), (11,28,27,20,16), (15,35,27,20,16)
5	Hexagon	(11,28,26,19,16,15), (11,28,27,20,16,15)
6	Heptagon	(11,28,27,26,19,16,15)
distance	polygon	$Solution-conv(52, distance, matrix, 1)$
3	Quadrilateral	(13,26,25,14)
4	Pentagon	(8,36,25,14,11)
5	Hexagon	(8,37,26,25,14,11)
6	Heptagon	(11,28,27,26,19,16,15)



(a) Triangle (b) Rectangle



(c) Quadrilateral simple - convex (d) Pentagon simple - convex

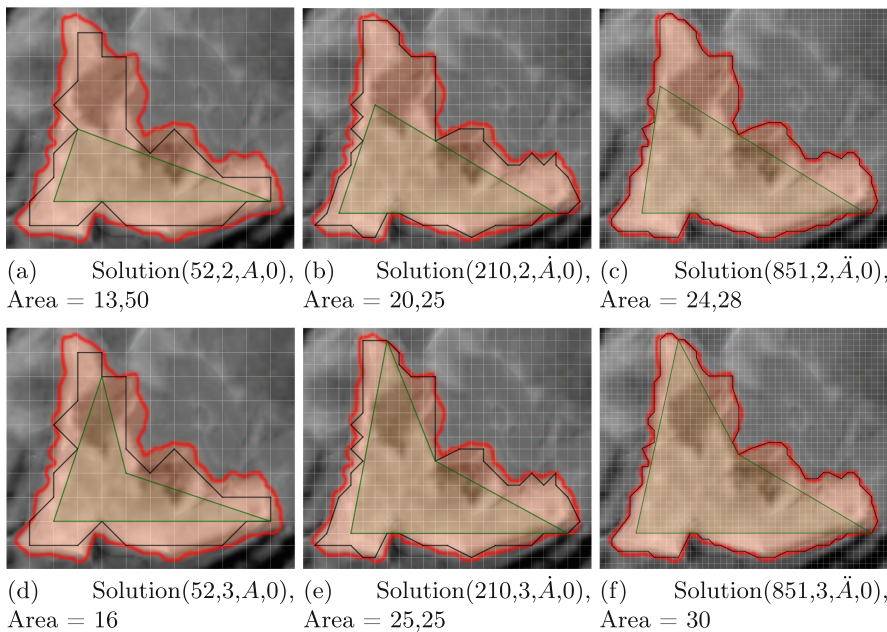


(e) Hexagon simple - convex (f) Heptagon simple - convex

**Fig. 3.** Example 1: solution for different polygons.

Figure 4 shows how to obtain the area of the closed contour  $C$ , by the inscribed limit area within the lattice polygon  $P$ , constructing finer partitions with  $L_{i+1} = L_i/2$  if  $\Pi_i \leq \Pi_{i+1}$  for all  $i$ . In this context, we present three partitions for the same closed contour  $C$ , with partition sizes  $L_1 = 1$ ,  $L_2 = 1/2$  and

$L_3 = 1/4$ , and number of points 52, 210 and 851, respectively, along with their adjacency matrix  $A$ ,  $\hat{A}$  and  $\ddot{A}$ . Using  $\text{Solution}(52,k,A,0)$ ,  $\text{Solution}(210,k,\hat{A},0)$  and  $\text{Solution}(851,k,\ddot{A},0)$ , we compute the maximum area simple  $k$ -sided polygon ( $k = 2, 3$ ). As shown in the solutions, the area of the simple  $k$ -sided polygon is larger if the partition is thinner each time.



**Fig. 4.** Maximum area simple  $k$ -sided polygon with different partition sizes ( $k = 2, 3$ ).

### 3.2 Photovoltaic Solar Energy Industry

From the application Visor SigPac v4.8 [1] we have obtained an image of a terrain partially covered by solar panels (Fig. 5a). The goal is to build more solar panels using the maximum possible area. Figure 5b shows the closed contour  $C$ , and Fig. 5c the lattice polygon  $P$  with  $N = \#(P) = 47$ . In addition, Fig. 6 displays the lattice polygon and Table 3 shows some of the solutions using Algorithm 2.

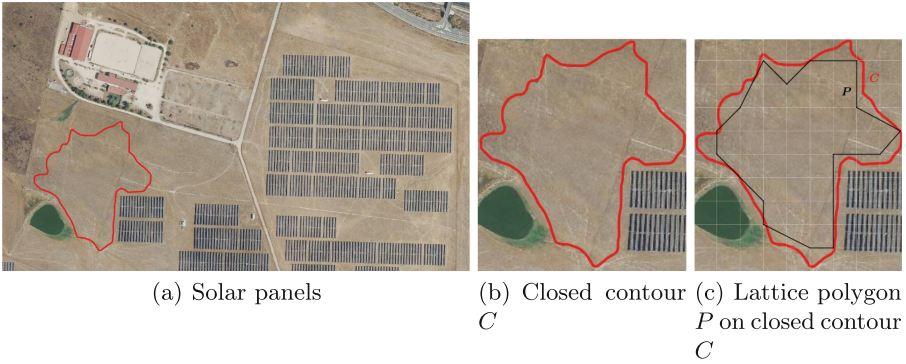


Fig. 5. Example 2: image extraction.

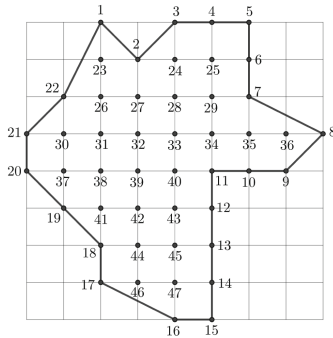


Fig. 6. Example 2: lattice polygon.

Table 3. Example 2: maximum-area simple and convex  $k$ -sided polygon.

distance	polygon	$Solution(47, distance, matrix, 0)$
2	Triangle	(5,21,16)
3	Quadrilateral	(3,20,16,5), (5,22,20,16)
	Rectangle	(14,17,23,25)
4	Pentagon	(1,20,16,5,2)
5	Hexagon	(1,20,16,5,3,2), (4,22,20,15,11,9), (4,22,20,16,11,9)
6	Heptagon	(4,22,20,16,15,11,9)
distance	polygon	$Solution-conv(47, distance, matrix, 0)$
3	Quadrilateral	(3,20,16,5), (5,22,20,16)
4	Pentagon	(4,22,20,16,5), (4,22,20,16,15)
5	Hexagon	(4,22,21,20,16,5), (4,22,21,20,16,15)
6	Heptagon	(2,22,21,20,16,15,25), (4,22,20,46,47,6,5), (4,22,21,20,46,47,6), (15,29,24,23,21,20,16)

## 4 Conclusions

The algorithm described in this paper determines the maximum area or perimeter of a simple  $k$ -sided polygon within a region of interest or closed contour. In order to calculate the solutions, the user must first select the type of polygon required by specifying a fixed value  $k$ , followed by selecting the desired solution: the maximum area or the maximum perimeter. Both types of solutions can also be computed when the polygon chosen is convex. Clearly, the algorithm is flexible since the user has the ability to decide what should be calculated. There is no other algorithm in the scientific literature that provides  $k$ -sided polygons of maximum area or maximum perimeter, inscribed in a closed contour.

There are numerous practical applications of the proposal, especially in the field of image processing and pattern recognition. In addition, the source code of the approach is available in several programming languages, so that anyone can use it in their scientific investigations.

## References

1. Visor sigpac v4.8 (2023). <https://sigpac.mapa.gob.es/fga/visor>
2. Alt, H., Hsu, D., Snoeyink, J.: Computing the largest inscribed isothetic rectangle. In: CCCG, pp. 67–72 (1995)
3. Boland, R.P., Urrutia, J.: Finding the largest axis-aligned rectangle in a polygon in... In: Proceedings of 13th Canadian Conference on Computing Geometry (2001)
4. van der Hoog, I., Keikha, V., Löffler, M., Mohades, A., Urhausen, J.: Maximum-area triangle in a convex polygon, revisited. *Inf. Process. Lett.* **161**, 105943 (2020)
5. Jin, K.: Maximal parallelograms in convex polygons, p. 376. arXiv preprint [arXiv:1512.03897](https://arxiv.org/abs/1512.03897) (2015)
6. Kallus, Y.: A linear-time algorithm for the maximum-area inscribed triangle in a convex polygon. arXiv preprint [arXiv:1706.03049](https://arxiv.org/abs/1706.03049) (2017)
7. Keikha, V.: Linear-time algorithms for largest inscribed quadrilateral (2020)
8. Knauer, C., Schlipf, L., Schmidt, J.M., Tiwary, H.R.: Largest inscribed rectangles in convex polygons. *J. Disc. Algor.* **13**, 78–85 (2012). <https://doi.org/10.1016/j.jda.2012.01.002>
9. Media Engineering Group (GIM): Source code, scripts, and documentation (2023). <https://github.com/UniversidadExtremadura/k-gon>
10. Melissaratos, E.A., Souvaine, D.L.: Shortest paths help solve geometric optimization problems in planar regions. *SIAM J. Comput.* **21**(4), 601–638 (1992). <https://doi.org/10.1137/0221038>
11. Molano, R., et al.: Finding the largest volume parallelepipedon of arbitrary orientation in a solid. *IEEE Access* **9**, 103600–103609 (2021). <https://doi.org/10.1109/ACCESS.2021.3098234>
12. Molano, R., Rodríguez, P.G., Caro, A., Durán, M.L.: Finding the largest area rectangle of arbitrary orientation in a closed contour. *Appl. Math. Comput.* **218**(19), 9866–9874 (2012). <https://doi.org/10.1016/j.amc.2012.03.063>
13. Rote, G.: The largest quadrilateral in a convex polygon, p. 381. *CoRR arxiv:1905.11203* (2019)





# Electricity Consumption Forecasting Using the Prophet Model in Industry: A Case Study

Umut Yildiz<sup>1</sup> (✉) and Sila Ovgu Korkut<sup>2</sup>

<sup>1</sup> Department of Software Engineering, Izmir Katip Celebi University, Izmir, Turkey  
umutyildiz@gmail.com

<sup>2</sup> Department of Engineering Sciences, Izmir Katip Celebi University, Izmir, Turkey

**Abstract.** Forecasting electricity consumption is a key mechanism for a wide range of industries to develop strategies or take precautions. This case study primarily aims to predict the electricity consumption of the machines on the production line through sensors and analyzers containing both thermocouples and devices storing the use of electricity. To do so, one of the most powerful methods, which is known for its ability to learn the main characteristics of the data for time-series models, the Prophet method has been utilized. Moreover, the capability of the Prophet method relying on both the use and not the use of temperature data in forecasting electricity consumption has been discussed. The achievement of the method has been supported through the tables and figures in both univariate and multivariate cases. Comparing the RMSE, MAE, and SMAPE scores, the results have shown that the not use of temperature data has been better than those of the use of temperature in the prediction.

**Keywords:** Electricity consumption · Prophet method · Forecasting

## 1 Introduction

Electricity consumption is a critical component of many industries, from manufacturing and retail to hospitality and healthcare. Forecasting electricity demand accurately is critical for businesses to optimize energy usage, costs, and sustainability. Therefore, generating any tool to predict electricity consumption has an increasing impact.

It is now well established from a variety of studies, that machine learning and deep learning algorithms as powerful tools for the estimation of electricity usage. Up to date, various machine learning algorithms have been studied to forecast electricity consumption: random forest model for short-term electricity load forecasting can be seen in [8], a long short-term memory (LSTM) recurrent neural network (RNN) based framework has been studied for short-term residential load forecasting in [11]. Convolutional neural networks (CNN) have been applied to predict solar power forecasting and electricity load forecasting in [12]. Ramos et al., [14] have compared the performance characteristics of artificial neural networks (ANN) and support vector machine (SVM) to forecast electricity consumption in an industrial facility for various data sets and learning parameters for different periods of the day. In addition to these models, several attempts have

been made to develop hybrid models. For instance, Chen et. al., [7], have proposed a new model based on hybridizing the fuzzy time series (FTS) and global harmony search algorithm (GHS) with least squares support vector machines (LSSVM) to make electric load forecasting. Another model, a direct optimum parallel hybrid (DOPH) model, based on multilayer perceptrons (MLP) neural network, Adaptive Network-based Fuzzy Inference System (ANFIS), and Seasonal Autoregressive Integrated Moving Average (SARIMA) has been proposed to electricity load forecasting in [5]. Bedi and Toshniwal, in [3], have developed a hybrid model using Empirical Mode Decomposition (EMD) and Long Short Term Memory (LSTM) network to improve the prediction accuracy of electricity demand forecasting models. The results of [3] highlighted that the EMD + LSTM outperformed the LSTM, the recurrent neural network (RNN), and the EMD + RNN for electricity demand time series forecasting. Besides the above-mentioned studies, a complete review of forecasting and improving the four main machine learning approaches including artificial neural network, support vector machine, Gaussian-based regressions, and clustering to has been presented in [16]. In [19] a new strategy has been suggested by integrating Long Short Term Memory (LSTM) models and Reinforcement Learning (RL) agents for forecasting building next-day electricity consumption and peak electricity demand. Moreover, a framework based on deep learning structures has been established in [4] to forecast electricity demand considering long-term historical dependencies. For further discussion on the machine learning and deep learning models for forecasting electricity consumption, we refer the interested reader to [9, 13, 15] and the references therein.

Whilst a great deal of attention have been paid on forecasting electricity consumption by various machine learning and deep learning methods, there have been a few investigations on the use of the Prophet model (ProphetAI) proposed in [18] for that target. Some of these studies can be exemplified as follows: The long-term maximum peak load forecasting performance of Kuwait has been investigated in [1] via the Prophet method compared to the Holt-Winters forecasting method. The Prophet model was reported to be preferable and robust over the Holt-Winter model by testing the methods on the real data extracted from National Control Center in Kuwait. Guo et al. [10] have studied the prediction of the maximum power demand (MPD) of industrial power users. To do so, once the related influencing factors determined by the improved Grey Relation Analysis (GRA), an improved MPD prediction algorithm based on the combination of the Fb Prophet method and Adaptive Cubature Kalman Filtering (ACKF) has been proposed according to the characteristics of customers' MPD. The goodness of performance of the proposed model is supported by experimental results in [10]. More recently, a hybrid method based on Prophet and LSTM models has been proposed by Bashir *et al.* to predict short-term electricity load. The results of the method in [2] were superior to those of standalone models like LSTM, ARIMA and etc. Moreover, Chaturvedi et al. have compared India's Central Energy Authority (CEA) existing trend-base modeling with SARIMA, LSTM RNN, and Fb Prophet in [6]. They concluded that performances of the Fb Prophet model and CEA model have better prediction errors than those in SARIMA, and LSTM RNN. Furthermore, Stefenon et al. [17] aggregated Prophet and seasonal trend decomposition using locally estimated scatterplot smoothing (LOESS) for forecasting Italian electricity spot prices.

To date, only a limited number of studies on forecasting electricity consumption in the industry via the ProphetAI have been identified. However, various advantages of using ProphetAI can be presented to anticipate industrial electricity use. One of the primary advantages is the improved prediction accuracy. ProphetAI can deliver more accurate estimates of future electricity demand by capturing complex patterns in data. This enables organizations to optimize their energy use, cut costs, and boost their sustainability. Additionally, saving time and resources can be counted as another advantage of ProphetAI. Current approaches for projecting power usages, such as regression or time series analysis, can be time-consuming and require significant skill. Model fitting and prediction, on the other hand, are automated with ProphetAI, saving businesses time and resources. Furthermore, ProphetAI can assist businesses in making more educated decisions about energy planning and sustainability. ProphetAI can assist businesses in assessing the risk associated with various energy usage scenarios by offering uncertainty estimates for its forecasts.

This study assesses the significance of the use of the ProphetAI on electricity consumption forecasting in industry, particularly the data of a factory producing terminal blocks, cable trays, and energy analyzers were used. Apart from the studies in the literature, using the prophet model; it has been observed whether the temperature values in certain parts of the molding machines affect the electricity consumption.

The current study is composed of three themed sections: Sect. 2 begins by describing the data collection and the processes to make the data ready for the Prophet model. It will then go on to provide the main architecture of the Prophet model. The results and discussions are presented in Sect. 3. A brief conclusion of this study will be given in Sect. 4.

## 2 Method and Materials

### 2.1 Data Collection and Preprocess

The goal of this research is to estimate how much electricity the machines use on the assembly line. To do so, in a factory where terminal blocks and electronic products are produced, data was collected from the machines on the production line through sensors and analyzers. Electricity consumption data were obtained by Klea v1.53 model devices. Thanks to the K-type thermocouples placed at certain points of the molding machine, the temperature data of the machine were collected. Then, temperature data every 15 min and total electricity consumption data every hour are added under the collections defined in MongoDB. The electricity consumption of the machines was stored in a database which uses MS SQL and MongoDB in JSON format in one-hour intervals. Once the data was gathered from the database, the data is composed of 63339 rows with five features (front-, mid-, back-, environmental temperatures as well as the electricity consumption in kilowatts per hour (kWh)) due to the sensors on the machine. Cleaning was done on the collected data. For example, if there is a null value in the values opposite the datetime arm due to communication problems, power cuts, that row is removed from the dataset. Likewise, the extreme values measured due to the error of the sensors in the dataset were also removed from the dataset. This process is followed by grouping data hourly by using mean of temperatures and sum of electricity consumption. Moreover,

the time scale of the data retained from the database was formatted in accordance with the algorithm. The final dataset has a shape of (5706,5).

## 2.2 Prophet Model (ProphetAI)

Taylor and Letham, from Facebook, suggested the Prophet model in [18] to generate modular configurable regression model taking into account the characteristics of the data. Looking at the series as curve fitting problems makes the ProphetAI more flexible, adaptable, interpretable. Taken together with its speed and accuracy the ProphetAI is more preferable to the other methods in many applications. ProphetAI is also a great instrument for projecting energy usage in the sector. Its capacity to handle complicated patterns in time-series data and estimate uncertainty makes it an appealing option for businesses trying to improve their energy usage and decrease costs. The mathematical model of the ProphetAI can be summarized as follows:

$$\hat{y}(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (1)$$

where  $\hat{y}(t)$  is the predicted value for time  $t$  and  $g(t)$  stands for the trend (non-periodic) component which is defined by

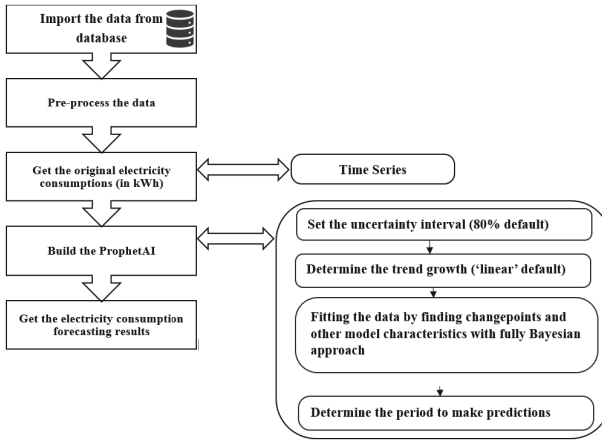
$$g(t) = \begin{cases} \frac{C(t)}{1 + \exp(-k_{adj}(t - m_{adj}))}, & \text{nonlinear, piecewise logistic growth} \\ k_{adj}t + m_{adj}, & \text{linear piecewise linear growth} \end{cases}$$

where  $k_{adj}$  and  $m_{adj}$  are the adjusted growth rate and adjusted offset parameters, respectively. Moreover,  $s(t)$  denotes the seasonality component, which is based on the Fourier series with a period, for capturing repeating patterns which is defined by

$$s(t) = a_1 \cos\left(\frac{2\pi(1)t}{P}\right) + b_1 \sin\left(\frac{2\pi(1)t}{P}\right) + \dots + a_N \cos\left(\frac{2\pi(N)t}{P}\right) + b_N \sin\left(\frac{2\pi(N)t}{P}\right)$$

where  $P$  stands for the regular period. Furthermore,  $h(t)$ , which returns the binary variable, represents the holiday component. It is defined by  $h(t) = \mathbf{W}(t)\boldsymbol{\kappa}$ , where  $\mathbf{W}(t) = [1(t \in D_1), \dots, 1(t \in D_L)]$  and  $\boldsymbol{\kappa} \sim \text{Normal}(0, \sigma_\kappa^2)$  where  $D_j$  represents the dates of the holidays for holiday  $j$ . Finally, the additive model has the irreducible residual component,  $\varepsilon_t$ . The ProphetAI employs a Bayesian optimization to find the optimal values for the hyper parameters, providing uncertainty intervals for the forecasts. For more details, we refer the interested reader to [18].

The pros and cons of the method can be summarized as follows: As more businesses embrace machine learning algorithms for forecasting, ProphetAI is likely to become a more popular option for projecting power consumption. However, it should be emphasized that estimating electricity usage with ProphetAI has significant limits. One of the primary disadvantages is that the prediction accuracy might be influenced by external factors that the model does not account for. Another disadvantage of utilizing ProphetAI is that it necessitates a large amount of data to train the model. While this is true for many machine learning algorithms, it is critical to have enough high-quality data to provide accurate forecasts. This can be difficult for some businesses that do not have access to adequate historical data or have data that is partial or untrustworthy.



**Fig. 1.** The architecture of the implementation to obtain the electricity consumption forecasting.

Before ending the section, the architecture of the implementation has been summarized in Fig. 1.

The explanations in this section present the use of the ProphetAI on the prediction of the electricity consumption of the machines. The next section, therefore, moves on to discuss the results the model predictions.

### 3 Results and Discussions

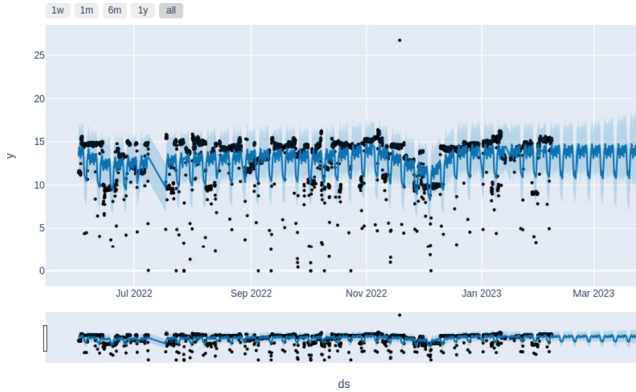
A factory producing electrical and electronic products; ProphetAI was used to estimate the electricity consumption of injection machines in the production line. In addition to the electricity consumption, the effect of temperature on the prediction of electricity consumption is also studied. The time span for this study is selected from 01/06/2022 to 04/04/2023. The descriptive statistics of the data has been summarized in Table 1.

**Table 1.** Summary of the dataset

	front	middle	back	environment	kwh
count	5706	5706	5706	5706	5706
mean	266.63	271.26	248.28	53.97	13.12
Standard deviation	16.26	16.77	19.06	21.92	2.46
min	17.82	18.63	17.66	0.01	0
max	301.56	303.98	276.72	77.78	26.72

After careful preprocessing, the shape of the final data set is 5706. Thereafter, the cleaned data is trained by ProphetAI determining the uncertainty interval as 80%. Moreover, the growth rate of the trend component is chosen linearly. The trained model is used

to make future predictions by selecting 1141 new data with hourly frequency in accordance with the structure of the dataset. The presented figure displays the projected and observed electricity consumption values, with forecasted values depicted by lines and real values by dots. The vertical axis of the plot represents the electricity consumption values, while the horizontal axis corresponds to the respective date values.



**Fig. 2.** Electricity consumption estimation using consumption values and date information

Figure 2 indicates how well ProphetAI is for prediction. To see how accurate predictions the ProphetAI provides, the performances of the ProphetAI with respect to the root mean square error (RMSE), mean absolute error (MAE), and symmetric mean absolute percentage error (SMAPE) has been discussed in Table 2.

$$\text{RMSE} : \sqrt{\frac{1}{N} \sum_{i=1}^N (y(t_i) - \hat{y}(t_i))^2} \quad (2)$$

$$\text{MAE} : \frac{1}{N} \sum_{i=1}^N |y(t_i) - \hat{y}(t_i)| \quad (3)$$

$$\text{SMAPE} : \frac{1}{N} \sum_{i=1}^N \frac{|y(t_i) - \hat{y}(t_i)|}{(|y(t_i)| + |\hat{y}(t_i)|)/2} \quad (4)$$

where  $y(t_i)$  and  $\hat{y}(t_i)$  stand for the actual value and the predictions, respectively.

While the prediction has been done hourly due to the structure of data, for the sake of simplicity three sample days were given in Table 2. Taking the listed metrics into account, Table 2 is evidence of the goodness of ProphetAI in predicting power. After performing 2 months of calculations, it was found that the highest root mean square error (RMSE) value observed was 6.0649 dated after the month, whereas the highest symmetric mean absolute percentage error (SMAPE) value was 60.0232 within the end of the second month, and at the end of the second month the maximum mean absolute error (MAE) value was 4.72. As seen by the low RMSE, MAE, and SMAPE values,

**Table 2.** Comparison of RMSE, MAE, and SMAPE

Date	RMSE	MAE	SMAPE(%)
2023-02-07	1.3931	1.3330	9.2043
2023-02-14	2.2503	2.2023	16.9684
2023-02-21	2.5764	2.5334	19.7708

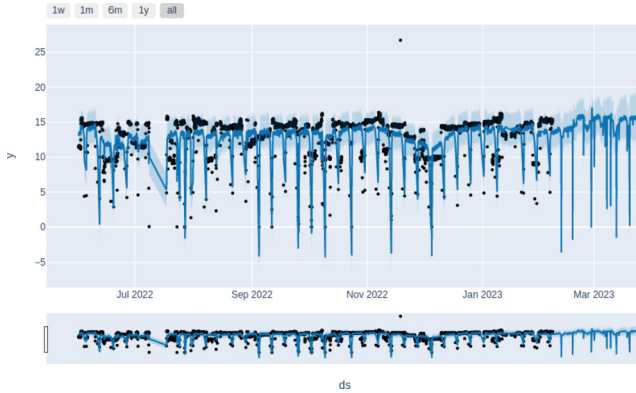
the results show that the machine learning model can accurately anticipate electricity usage on the majority of dates. The performance of the model, however, may change depending on the precise date and outside variables that affect electricity use. These results imply that the ProphetAI can be an effective tool for companies to control their energy use and lower expenses, but it should be used with caution and in conjunction with other forecasting techniques.

Another application was designed to gauge how much effective the temperature is in the forecasting of electricity consumption via ProphetAI. After the Prophet model was created, temperature values were added as an external regression element. And the model was run again. Figure 3 illustrates the electricity consumption forecasting via ProphetAI taking into account effects of temperature. The graphical representation in Fig. 3 illustrates the predicted and actual values of electricity consumption. Similar to those given in the case of the future electricity consumption prediction using just date and consumption values, three sample days were given in Table 3 under three metrics, namely RMSE, MAE, and SMAPE.

**Table 3.** The accuracy analysis of the prediction of the ProphetAI considering the temperature effects

Date	RMSE	MAE	SMAPE
2023-02-07	1.6398	1.6254	11.2986
2023-02-14	2.1570	2.1468	16.6238
2023-02-21	4.2719	4.2649	31.2520

The effectiveness of ProphetAI model in forecasting power use for three different days is shown in the table above. The outcomes of the 2-month calculations revealed that the maximum root mean squared error (RMSE) value was 7.9932 dated at the end of the second week, indicating a relatively large discrepancy between predicted and actual values. The maximum symmetric mean absolute percentage error (SMAPE) value was found at the eighth week to be 138.7083, indicating a considerable level of inaccuracy in the predictions. Additionally, the maximum mean absolute error (MAE) value was 7.2600 by the second week, indicating a relatively high level of error in the predicted values. These results have significant implications for companies and organizations that depend on precise electricity consumption forecasts to control their energy expenses and usage.



**Fig. 3.** Electricity consumption estimation using Kwh, temperature value of the sensor located at the back of the machine information, and date information. The predicted values are displayed as lines, while the actual values are represented by dots. Notably, the estimation includes temperature values as an input feature. Values in the y-plane show electricity consumption values.

## 4 Conclusion

In conclusion, accurate predictions are essential for managing energy usage, cutting costs, and enhancing sustainability. Estimating power consumption is a fundamental component of many sectors. Given their capacity to recognize intricate patterns in time-series data and offer uncertainty estimates, machine learning algorithms like ProphetAI have become an effective tool for forecasting electricity usage. This study was designed to assess the use of ProphetAI for forecasting the electricity consumption of machines in industry. To that aim, data was collected from the machines on the production line. Two different subsets of data were considered. One of the subsets was date and the electricity consumption and the other was the use of date, the electricity consumption as well as the temperature information. In the end, the model proved to be an effective substitute for traditional methods of estimating electricity use. Although the user-friendly interface and streamlined development process of the model are important benefits, the correct dataset preparation is still a key factor to take into account. Notably, the results of this investigation show that the performance of the model's estimation is adversely affected by the addition of each external regression feature. Besides, the not use of temperature gives better RMSE, MAE, and SMAPE records. The insights gained from this study may be of assistance to increase the impact of the use of ProphetAI in industry. Moreover, one can be inferred from this study is the combination of ProphetAI with various models which is opening an area for future work that can make significant contributions.

**Acknowledgement.** This work was supported by the Klemsan Elektrik Elektronik A.Ş.



## References

1. Almazrouee, A.I., Almeshal, A.M., Almutairi, A.S., Alenezi, M.R., Alhajeri, S.N.: Long-term forecasting of electrical loads in Kuwait using prophet and holt-winters models. *Appl. Sci.* **10**(16), 5627 (2020). <https://doi.org/10.3390/app10165627>
2. Bashir, T., Haoyong, C., Tahir, M.F., Liqiang, Z.: Short-term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN. *Energy Rep.* **8**, 1678–1686 (2022). <https://doi.org/10.1016/j.egy.2021.12.067>
3. Bedi, J., Toshniwal, D.: Empirical mode decomposition based deep learning for electricity demand forecasting. *IEEE Access* **6**, 49144–49156 (2018). <https://doi.org/10.1109/ACCESS.2018.2867681>
4. Bedi, J., Toshniwal, D.: Deep learning framework to forecast electricity demand. *Appl. Energy* **238**, 1312–1326 (2019)
5. Chahkoutahi, F., Khashei, M.: A seasonal direct optimal hybrid model of computational intelligence and soft computing techniques for electricity load forecasting. *Energy* **140**, 988–1004 (2017). <https://doi.org/10.1016/j.energy.2017.09.009>
6. Chaturvedi, S., Rajasekar, E., Natarajan, S., McCullen, N.: A comparative assessment of SARIMA, LSTM RNN, and Fb Prophet models to forecast total and peak monthly energy demand for India. *Energy Policy* **168**, 113097 (2022). <https://doi.org/10.1016/j.enpol.2022.113097>
7. Chen, Y.H., Hong, W.-C., Shen, W., Huang, N.N.: Electric load forecasting based on a least squares support vector machine with fuzzy time series and global harmony search algorithm. *Energies* **9**(2), 70 (2016). <https://doi.org/10.3390/en9020070>
8. Dudek, G.: Short-term load forecasting using random forests. In: Filev, D., Jabłkowski, J., Kacprzyk, J., Krawczak, M., Popchev, I., Rutkowski, L., Sgurev, V., Sotirova, E., Szykarczyk, P., Zadrozny, S. (eds.) *Intelligent Systems'2014*. AISC, vol. 323, pp. 821–828. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-11310-4\\_71](https://doi.org/10.1007/978-3-319-11310-4_71)
9. González-Briones, A., Hernandez, G., Corchado, J.M., Omatu, S., Mohamad, M.S.: Machine learning models for electricity consumption forecasting: a review. In: *2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, pp. 1–6. IEEE (2019)
10. Guo, C., Ge, Q., Jiang, H., Yao, G., Hua, Q.: Maximum power demand prediction using fbprophet with adaptive Kalman filtering. *IEEE Access* **8**, 19236–19247 (2020). <https://doi.org/10.1109/ACCESS.2020.2968101>
11. Kong, W., Dong, Z.Y., Jia, Y., Hill, D.J., Xu, Y., Zhang, Y.: Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Trans. Smart Grid* **10**(1), 841–851 (2019). <https://doi.org/10.1109/TSG.2017.2753802>
12. Koprinska, I., Wu D., Wang, Z.: Convolutional neural networks for energy time series forecasting. In: *2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, Brazil, pp. 1–8 (2018). <https://doi.org/10.1109/IJCNN.2018.8489399>
13. Ozcanli, A.K., Yaprakdal, F., Baysal, M.: Deep learning methods and applications for electrical power systems: a comprehensive review. *Int. J. Energy Res.* **44**, 7136–7157 (2020). <https://doi.org/10.1002/er.5331>
14. Ramos, D., Faria, P., Vale, Z., Correia, R.: Short time electricity consumption forecast in an industry facility. *IEEE Trans. Ind. Appl.* **58**(1), 123–130 (2022). <https://doi.org/10.1109/TIA.2021.3123103>
15. Saurabh, V., Vaibhav, K., Mrunmayee, G., Atharva, C., Sandip, H.: A survey: electricity demand prediction using statistical, machine learning and deep learning methods. *Int. Res. J. Modern. Eng. Technol. Sci.* **05**(02), 493–502 (2023). <https://doi.org/10.56726/IRJMET.533460>

16. Seyedzadeh, S., Rahimian, F., Glesk, I., et al.: Machine learning for estimation of building energy consumption and performance: a review. *Vis. Eng.* **6**, 5 (2018). <https://doi.org/10.1186/s40327-018-0064-7>
17. Stefenon, S.F., Seman, L.O., Mariani, V.C., Coelho, L.D.S.: Aggregating prophet and seasonal trend decomposition for time series forecasting of Italian electricity spot prices. *Energies* **16**(3), 1371 (2023). <https://doi.org/10.3390/en16031371>
18. Taylor, S.J., Letham, B.: Forecasting at scale. *Am. Stat.* **72**(1), 37–45 (2018)
19. Zhou, X., Lin, W., Kumar, R., Cui, P., Ma, Z.: A data-driven strategy using long short term memory models and reinforcement learning to predict building electricity consumption. *Appl. Energy* **306**, 18078 (2022). <https://doi.org/10.1016/j.apenergy.2021.118078>



# Going Beyond Traditional Methods: Using LSTM Networks to Predict Rainfall in Kerala

J. Akshaya<sup>1</sup>, D. Harsha<sup>1</sup>, D. Eswar Chowdary<sup>1</sup>, B. E. Pranav Kumar<sup>1</sup>,  
G. Rahul<sup>1</sup>, V. Sowmya<sup>1</sup>(✉), E. A. Gopalakrishnan<sup>1,2</sup>, and M. Dhanya<sup>3</sup>

<sup>1</sup> Centre for Computational Engineering and Networking (CEN),  
Amrita Vishwa Vidyapeetham, Coimbatore, India  
v\_sowmya@cb.amrita.edu

<sup>2</sup> Department of Computer Science and Engineering, Amrita School of Computing,  
Amrita Vishwa Vidyapeetham, Bangalore, India

<sup>3</sup> Center for Wireless Networks and Applications (WNA), Amrita Vishwa  
Vidyapeetham, Amritapuri, India

**Abstract.** India has witnessed a notable upsurge in floods owing to shifts in global climatic patterns underlining climate change and global warming. This inorganic change is significantly evident in places like Kerala and Tamil Nadu. Kerala receives perennial rainfall throughout the year, influenced by southeast and northwest monsoon rainfall cycles, resulting in flood-induced calamities like landslides prevailing in the Thrissur and Munnar regions of Kerala. Long-Term Short Memory (LSTM) networks outperform pre-existing deep learning forecasting models as it is optimal for time series forecasting by handling non-linear spatiotemporal dynamics, and adapting to long-term dependencies in time series data while ensuring high scalability. The objective of the present work is rainfall prediction in flood-prone regions in Kerala. We propose an LSTM network to predict the monthly rainfall in Thrissur, Pathanamthitta, Munnar and Kottayam. The proposed model is evaluated using the metrics like mean absolute error and root-mean-squared error.

**Keywords:** Rainfall Forecasting · Climate Change · Time Series · LSTM

## 1 Introduction

Accurate and reliable rainfall prediction is very crucial for the various sectors of our society. It has wide-ranging implications across multiple domains, such as agriculture, disaster management, water management, infrastructure management, and tourism. India, with its agrarian roots, could greatly thrive from real-time scalable rainfall forecasts. Unique monsoon patterns which are found in states like Kerala make it challenging to create and conceive a high-fidelity robust model which can accommodate such conditions and shifting time and space dynamics.

Hong et al. proposed a hybrid model for rainfall prediction with Recurrent Artificial Neural Networks (RNN), Support Vector Machine (SVM), and Chaotic Particle Swarm Optimization (CSPO) Algorithm [1]. The SVM mapped the data into a higher feature space and CSPO optimized the feature selection process. The selected features were used by an RNN for further predictions. The model predicted peak occurrences but its accuracy was low, especially at the peaks.

Sumi et al. proposed a hybrid multi-model forecasting technique for rainfall prediction by comparing various algorithms in input selection, data pre-processing, and forecasting models [2]. They found Linear Discriminant Analysis (LDA) to be the most effective input selection technique, Principle Component Analysis (PCA) to be efficient in data pre-processing, and SVR to perform better than other models in forecasting.

Emily et al. proposed a deep learning architecture comprising an autoencoder network and a multilayer perceptron (MLP) network for predicting daily precipitation [3]. The autoencoder network extracts non-linear features for data input, while the MLP uses one hidden layer and the sigmoid activation function for classification and prediction tasks. The proposed model outperformed earlier-designed models in terms of Root-Mean-Squared-Error (RMSE) and Mean Squared Error (MSE).

Sam et al. proposed a study to compare seven machine learning algorithms, such as k-Nearest Neighbours, Genetic Programming, M5 Rules, Support Vector Regression, M5 Model trees, Radial Basis Neural Networks, and the current Markov chain extended with rainfall prediction (MCRP), for rainfall prediction [4]. They used two decades of daily statistics and found that drier or moister weather conditions were correlated with lower predictive errors, while highly volatile cities and extreme monsoon seasons were associated with higher predictive errors.

Risul Islam et. proposed Support Vector Regression (SVR) and Artificial Neural Networks (ANN) based model using a 6-year historical temperature and precipitation intensities of the Chittagong metropolitan area [5]. The study found that SVR had a marginal error rate and outperformed ANN in rainfall prediction, while ANN produced accurate predictions as compared to SVR with an acceptable error rate deviation.

Barrera-Animas et al. proposed a comparative study on deep learning architectures and traditional machine learning algorithms for forecasting hourly rainfall volumes [6]. Using climate data from five major UK cities, models based on Long-Short Term Memory (LSTM), XGBoost, Stacked-LSTM, an ensemble of Gradient Boosting Regressor, Bidirectional-LSTM Networks, Linear Support Vector Regression, and an Extra-trees Regressor were evaluated. Results showed that LSTM networks with a high number of hidden layers are less effective in forecasting hourly rainfall values.

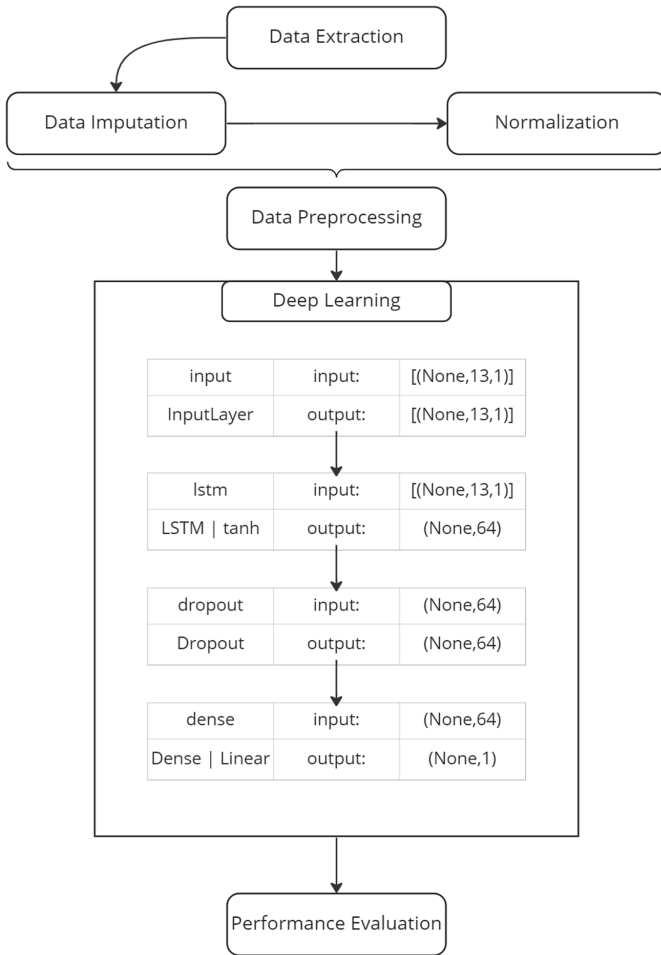
Balamurugam et al. proposed a clustering algorithm based on environmental conditions like temperature and humidity to predict the probability of rainfall in a particular region [7]. A network setup similar to a wireless sensor network was used for feature extraction. Results showed that machine learning models have

a minimal error in short-term rain forecasting, with only 2-day rain forecasting being reliable.

This work is structured as follows: the proposed methodology is described in Sect. 2, and results and discussion are presented in Sect. 3. The conclusion can be found in Sect. 4.

## 2 Methodology

This paper proposes an approach that is illustrated in the flowchart shown in Fig. 1. The flowchart outlines the step-by-step procedure that is followed in this approach.



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

## 2.1 Data Extraction

Data extraction is crucial for constructing advanced prediction models. The daily rainfall data provided by the India Meteorological Department (IMD) is utilized in the current study. IMD creates a high-quality gridded dataset of horizontal resolution  $0.25 \times 0.25^\circ$  incorporating observations from over 6000 rain gauges across the country [8].

The dataset is available from (<https://imdpune.gov.in/lrfindex.php>). IMD data is available from 1901 onwards, arranged in  $135 \times 129$  grid points. For this study, the daily rainfall values for the grid points over Kerala state were extracted. The focus is on 4 primary grid points with distinct and notably useful rainfall prediction outcomes. The points corresponding to Munnar, Pathanamthitta, Kottayam, and Thrissur were shortlisted based on a preliminary exploratory study of historic correlated events [9].

## 2.2 Data Preprocessing

Data preprocessing is an indispensable part of the data preparation process. It is essential for avoiding performance issues.

**Normalization.** The data can now be reduced to a convenient format using the process of normalization. In addition, the data also becomes more relatively interpretable for the model improving its overall performance. The MinMaxScaler is a simple and effective normalization method that scales the data to  $[0, 1]$ . Mathematically it individually operates on each data measurement as shown in the following Eq. 1, where  $x_{min}$  and  $x_{max}$  correspond to the minimum and maximum value of  $x$ ,

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

## 2.3 Model

Deciding the model architecture is a critical part of the design process which must be backed by experimental data and research specific to the task at hand. The contemporary deep Learning approach is to solve time-series data-based prediction tasks using LSTM. The model should capture instantaneous and prolonged innate patterns adequately.

Experiments were conducted on different layer depths and widths in an LSTM model to capture pseudo-periodic data while avoiding high-detail noise. It was found that models with deeper layers showed poor performance. Layer width was found to determine the level of detail captured, and 64 was chosen as an arbitrary value. The optimal window size for capturing seasonal weather patterns was found to be 13, striking a balance between providing relevant information and avoiding irrelevant data. These findings were based on experiments that aimed to maximize model performance by retaining crucial marker amplitudes while avoiding noise.

## 2.4 LSTM

Long-term dependencies can be learned by LSTM models. The LSTM requires four elements rather than a single neural network layer, like an RNN. Cells ( $\tanh$ ), an input gate ( $I_t$ ), an output gate ( $O_t$ ), and a forget gate ( $\Gamma_t$ ) make up a hidden layer. Cells store the values of random time intervals, and three gates control the information flow into and out of cells. These gates are mainly employed to regulate the data to be stored. The LSTM network implements temporary storage by switching the gate in order to prevent vanishing gradients.

$$\Gamma_t = \sigma(\omega_f \cdot [\kappa_{t-1}, x_t] + \beta_f) \quad (2)$$

$$I_t = \sigma(\omega_i \cdot [\kappa_{t-1}, x_t] + \beta_i) \quad (3)$$

$$O_t = \sigma(\omega_o \cdot [\kappa_{t-1}, x_t] + \beta_o) \quad (4)$$

$$C_t = \tanh(\omega_C \cdot [\kappa_{t-1}, x_t] + \beta) \quad (5)$$

LSTM models can efficiently process input time series data due to the horizontal stacking of cells without compatibility issues. The retained favorable properties of the cells enhance adaptability and flexibility, making LSTM models suitable for various applications.

## 2.5 Model Description

Our model was developed using the Keras API of TensorFlow [10] and follows the architecture shown in Fig. 1. We experimented with numerous configurations by varying hyper parameters ranging from 1–5 layers, 32–512 units (in multiples of 2), 5–50, 0.0–0.4, and  $10^{-4}$ – $10^{-1}$  for LSTM layers, per layer units, window sizes, dropout rate, and learning rate respectively. Finally, via empirical adjustments, we found that the model performs best with the following configuration which includes a single LSTM layer with 64 units, a model window size of [12, 13, 14], and a dropout rate of [0.0, 0.2]. The model was trained using the adam optimizer with a learning rate of 0.005 for 50 epochs and a batch size of 32. The training data spans from 1992 to 2016, while the testing data is from 2017 to 2018.

## 2.6 Performance Metrics

For evaluating the rainfall forecasting performance of our deep learning model we used the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

The Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are the two most common performance metrics used in continuous problems. MAE measures the average magnitude of the errors in a set of predictions, without considering their direction.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

Here,  $y_i$  is the actual value, and  $\hat{y}_i$  is the predicted value, where  $n$  is the number of data points.

The Root Mean Squared Error (RMSE) is a quadratic scoring rule that also measures the average magnitude of the error. It is the square root of the average of squared differences between prediction and actual Values.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

The Root Mean Squared Error (RMSE) places a greater role in emphasizing larger errors than smaller ones. As a result, it is a more appropriate metric when the impact of large errors is significant and should be weighted more heavily. On other hand, we have Mean Absolute Error (MAE) to analyze the error in the model.

### 3 Experimental Results and Discussion

Our experimentation process is a continuous approach that begins with the procurement of data and cleaning of the rainfall dataset. We then proceed to train the model and make forecasts. To validate the performance of the model, we benchmark the model using various performance measures. This ensures that our trained model is of the best quality and provides better forecasts. By pursuing this approach, we can ensure that our model is effective in predicting rainfall for our four selected locations in Kerala, thus providing beneficial information for an early flood warning system in those areas that are vulnerable to landslides and other weather-related hazards.

After analyzing the data on floods and landslides, we chose Munnar, Kottayam, Pathanamthitta, and Thrissur as the locations for our rainfall forecasting model in Kerala. The training set for the model spanned from 1992 to 2016, while the evaluation set was from 2017 to 2018. This data range was chosen for optimal performance, as the model accurately captured seasonal rainfall patterns in the region.

Before discussing the results, let's review some attributes of the rainfall forecasting model. The window size determines how many past time steps are used as input to predict the following time period. The learning rate is the size of the steps taken to update the model weights. The input layer size in the LSTM is the number of neurons in the input layer. Dropout refers to the fraction of input units that are dropped. Dropout is to mitigate overfitting in the model. Table 1 gives the attributes used in the proposed model.

The reason for having multiple window sizes is to identify the distinct patterns and trends in the data for each location, various window sizes may be employed for each LSTM model. The optimal window size will depend on the



available time series data, the complexity of the rainfall patterns, and the temporal resolution of the data.

The longer window size can help in capturing long-term dependencies in the data, but it can also increase the likelihood of overfitting. On the other hand, we have a shorter window size which can accurately capture short term dependencies in the data but it can also lead to underfitting of the model, but it might not be suitable for capturing long-term patterns in the data.

**Table 1.** Hyperparameters of the proposed model

Learning rate	Optimizer	Input layer	Batch Size	Epochs
0.005	Adam	64	32	50

**Table 2.** Performance metrics of the proposed LSTM model for various districts

District	Window Size	Dropout	MAE	RMSE
Munnar	12	0.0	4.73	7.13
Munnar	<b>13</b>	<b>0.0</b>	<b>3.86</b>	<b>5.38</b>
Munnar	14	0.0	6.23	6.76
Munnar	12	0.2	4.69	6.71
Munnar	13	0.2	5.62	8.13
Munnar	14	0.2	5.94	8.86
Pathanamthitta	12	0.0	5.11	6.18
Pathanamthitta	13	0.0	5.03	5.08
Pathanamthitta	14	0.0	5.34	6.39
Pathanamthitta	12	0.2	5.05	6.18
Pathanamthitta	<b>13</b>	<b>0.2</b>	<b>4.01</b>	<b>5.24</b>
Pathanamthitta	14	0.2	5.19	5.19
Kottayam	12	0.0	7.49	8.86
Kottayam	13	0.0	5.88	7.56
Kottayam	<b>14</b>	<b>0.0</b>	<b>5.33</b>	<b>6.31</b>
Kottayam	12	0.2	5.03	6.86
Kottayam	13	0.2	7.06	8.87
Kottayam	14	0.2	5.52	6.68
Thrissur	12	0.0	7.66	9.98
Thrissur	13	0.0	4.51	5.73
Thrissur	<b>14</b>	<b>0.0</b>	<b>4.64</b>	<b>5.17</b>
Thrissur	12	0.2	5.03	6.86
Thrissur	13	0.2	7.35	8.70
Thrissur	14	0.2	6.19	8.08

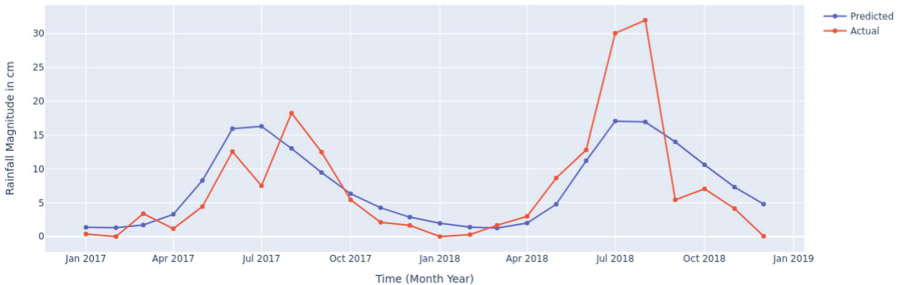
Thus, each LSTM model’s window size should be carefully chosen based on the specific characteristics of the data for each location. The goal is to find a balance between capturing short-term and long-term patterns in the data while avoiding overfitting. Table 2 summarizes the results of the rainfall forecasting model for the four selected locations in Kerala.

We used the results from Table 2 to select the best model for each location. We then generated line plots comparing the forecasted and actual rainfall for the years 2017 and 2018. These plots, given in Figs. 2, 3, 4, and 5, have time on the x-axis and rainfall in centimetres on the y-axis.

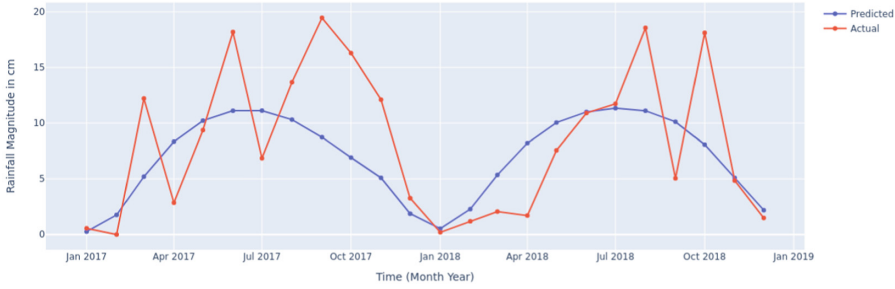
The results of our experiments have been documented above in tabular form 2, detailing the hyperparameters applied with the corresponding RMSE and MAE errors. The graphs from Figs. 2, 3, 4, and 5 illustrate the degree of optimal fit for all four locations under consideration.

We can see that regions like Munnar and Kottayam displayed significantly superior results compared to Pathanamthitta and Thrissur. The discrepancies in our projections are suspected due to climatic irregularities and anomalies that have occurred across the loaded month-wise averaged input data. Kerala witnessed one of the worst droughts in the century in 2016 and 2017 along with catastrophic floods in 1999 and 2018, the former being dubbed as the ‘Great flood of 99’.

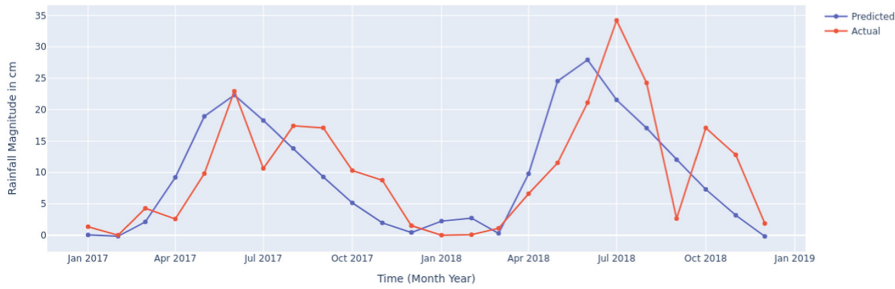
To maximize the run time efficiency, the adam optimizer is selected over other standard optimizers like stochastic gradient descent. To minimize overfitting, the number of epochs the model trained is upon is limited for each geographical location. This promotes capturing only the salient features of our training data and predicting the succeeding rainfall estimate for the next time step. The learning rate, which determines the extent to which historical data is implicitly considered from the previous step, has been thoroughly scrutinized. Upon investigation, we discovered the ideal number of epochs as fifty and the learning rate to be 0 or 0.2 for various locations. The cumulative collection of these determined hyperparameters produces the best results for our model.



**Fig. 2.** Rainfall forecasting for Munnar from 2017 to 2018



**Fig. 3.** Rainfall forecasting for Pathanamthitta from 2017 to 2018



**Fig. 4.** Rainfall forecasting for Kottayam from 2017 to 2018



**Fig. 5.** Rainfall forecasting for Thrissur from 2017 to 2018

## 4 Conclusion

The proposed LSTM neural network is a powerful tool for rainfall prediction, which accounts for floods in the past decade and provides results for real-world applications. By leveraging its ability to capture spatiotemporal dependencies in data, LSTM can effectively forecast peaks in rainfall, which represent an unusually high amount of rainfall. This enables early warning triggers for potential flood events. Overall, the promising performance of LSTM in rainfall prediction highlights its potential to improve disaster risk reduction efforts and enhance

resilience in vulnerable communities. Future work can focus on enhancing the predictive capabilities of LSTM (and other models) by incorporating additional features or data sources. It can also explore the application of other models such as GRU, BiLSTM, etc. to further enhance forecasting accuracy and performance. Moreover, exploring ensemble techniques, which combine the predictions of multiple models, can potentially improve the robustness of rainfall forecasting systems.

## References

1. Hong, W.-C.: Rainfall forecasting by technological machine learning models. *Appl. Math. Comput.* **200**(1), 41–57 (2008). ISSN 0096-3003
2. Monira Sumi, S., Faisal Zaman, M., Hirose, H.: A rainfall forecasting method using machine learning models and its application to the Fukuoka city case. *Int. J. Appl. Math. Comput. Sci.* **22**(4), 841–854 (2012)
3. Hernández, E., Sanchez-Anguix, V., Julian, V., Palanca, J., Duque, N.: Rainfall prediction: a deep learning approach. In: Martínez-Álvarez, F., Troncoso, A., Quintián, H., Corchado, E. (eds.) *HAIS 2016*. LNCS, vol. 9648, pp. 151–162. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-32034-2\\_13](https://doi.org/10.1007/978-3-319-32034-2_13)
4. Cramer, S., et al.: An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. *Expert Syst. Appl.* **85**, 169–181 (2017). ISSN 0957-4174
5. Rasel, R.I., Sultana, N., Meesad, P.: An application of data mining and machine learning for weather forecasting. In: Meesad, P., Sodsee, S., Unger, H. (eds.) *IC2IT 2017*. AISC, vol. 566, pp. 169–178. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-60663-7\\_16](https://doi.org/10.1007/978-3-319-60663-7_16)
6. Barrera-Animas, A.Y., et al.: Rainfall prediction: a comparative analysis of modern machine learning algorithms for time-series forecasting. *Mach. Learn. Appl.* **7**, 100204 (2022). ISSN 2666-8270
7. Balamurugan, M.S., Manojkumar, R.: Study of short term rain forecasting using machine learning based approach. *Wirel. Netw.* **27**(8), 5429–5434 (2021). ISSN 1572-8196
8. Pai, D.S., et al.: Development of a new high spatial resolution ( $0.25^\circ \times 0.25^\circ$ ) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *MAUSAM* **65**(1), 1–18 (2014)
9. Flood - KSDMA. <https://sdma.kerala.gov.in/flood/>. Accessed 07 Mar 2023
10. Abadi, M., et al.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems (2015). Software available from <http://tensorflow.org/>



# Effects of Augmented Reality on Visuospatial Abilities of Males and Females

Julia Bend<sup>(✉)</sup> and Anssi Öörni

Åbo Akademi University, Turku, Finland

julia.bend@abo.fi

**Abstract.** Males and females inherently perceive cognitive load differently due to their biological make-up. Research indicates that males outperform females in spatial tasks. There is an increased demand for solutions to minimise the gap in visuospatial abilities between sexes. Augmented reality (AR) techniques offer a range of options to minimise the inter-sex gap in visuospatial perception and cognitive processes. However, research on the causes of cognitive differences between sexes in visuospatial tasks is obscure. Studies have shown that males have better reaction time and accuracy than females when performing spatial visualisation and orientation tasks with AR. There are no investigations on the factor of sex-specific differences that might impact user performance. Hence, this research focuses on inter-sex differences in perceptions of AR solutions and their effect on user reaction time and accuracy. The study employs an alternative research design with spatial AR – involving object-identification processing and spatial processing – to research the design criteria for AR solutions. The goal is to reduce participants’ cognitive load and reaction time and, ultimately, increase their performance.

**Keywords:** AR · augmented reality · visuospatial skills · sex differences

## 1 Introduction

Differences in cognition and somatosensation (sensations of temperature, pain, balance, vibration, discriminatory touch and pressure, positioning and movement) attributable to sex are well-established [28]. Contrarily, less attention has been dedicated to sex-specific differences in basic visual sensations determined at “early” stages of the eye and brain [1]. Males tend to outperform females in tasks that involve visual perception [28]. Furthermore, “far-vision” and “near-vision” differ significantly between sexes, as men generally detect and assess the sizes of distant targets with better precision [29]. According to Newcombe [25], there is an acknowledged gap in visuospatial ability between sexes – males outperform females in spatial tasks. In general, visuospatial ability is the capacity to represent, transform, generate and recall symbolic, non-linguistic information [22]. This ability includes several domains, viz.: “visualisation”, “spatial relations” and “orientation”, which implies that skills in orientation and the mental manipulation of visual stimuli are both recognised under the same skill [13]. The above-mentioned inter-sex

differences may result in the underrepresentation of women in STEM (Science, Technology, Engineering and Mathematics) fields due to the importance of visuospatial and mathematical skills – the latter of which is thought to be mediated by visuospatial abilities [7, 8, 25].

Therefore, there is currently an increased demand for finding ways to narrow the inter-sex differences in visuospatial abilities. For example, spatial practice and training, among other options, can be applied as a strategy to reduce this disparity [8, 24]. However, it is uncertain whether dispersed practice may provide meaningful and useful gains. It is also unclear how variations in personality impact practice results. In contrast, this study examines how the use of technology, specifically augmented reality (AR – overlaying virtual objects onto the real world), could enhance individual performance in visuospatial tasks.

## 2 Related Work

### 2.1 Visuospatial Tasks and Different Cortical Activation Patterns

The largest sex-specific difference in visuospatial tasks is accessed in mental rotation tasks – where males outperform females [8, 22, 26]. The differences in strategies between sexes in mental rotation tasks result in less efficient verbal and analytical strategies and lower test performances for females [6, 15, 21]. However, the causes of inter-sex cognitive differences in visuospatial tasks are obscure. A number of brain imaging studies have examined the regional activation of the brain during mental rotation tasks [11, 18, 30, 33] and some of them addressed inter-sex differences. Interestingly, Thomsen et al. [31] found significantly different activation patterns in females and males. Males showed predominantly parietal activation; in contrast, females showed inferior frontal activation [31]. Jordan et al. [18] found that the superior parietal lobule, premotor and pre-supplementary motor areas for males and females were evoked by mental rotation tasks; however, the stronger activations of inferior temporal gyrus (ITG) and middle occipital gyrus (MOG) was found only for females. According to Carpenter et al. [4], the activation of ITG and MOG connects with an object and object-part identification and the “what” system. Hence, there are sex-specific preferred strategies for solving cognitive tasks and various activations patterns in the brain. Interestingly, the amount of activation in the different cortical regions is related to the amount and type of computational demand during visuospatial tasks [4]. Therefore, one of the possible interpretations of stronger activations in ITG could be that females place a greater demand on the completion of visuospatial tasks than males, resulting in higher cognitive load.

### 2.2 Sex Differences in Perception of AR

AR techniques that enhance the user’s environment with computer-generated perceptual information in real-time offer an interesting range of options to narrow the inter-sex gap in visuospatial perception and cognitive processes.

Recent studies [12] show that female participants are more enthusiastic about using new technologies, and AR solutions trigger more positive emotions and higher engagement than in males. According to Goswami and Dutta [16], sex has a significant impact

on the commitment to adopt new technology. Owing to inherent biological and cognitive differences, males and females perceive cognitive load differently [38]. Previous studies suggest that AR is more effective for novice users who experience considerable cognitive load; however, there are no investigations on the factor of sex differences that might impact user performance [17]. Unfortunately, research on inter-sex differences in the use of AR is scarce [12]. Ahmad et al. [2] indicate that males perform better than females in spatial visualisation and orientation tasks with AR, both in reaction time and accuracy. Males also outperformed females in the tests by Waller et al. [36] involving navigation.

Many studies [8, 13, 14, 24] observed spatial practice and training as a strategy to improve visuospatial abilities. Durlach et al. [14] claim that it is possible to strengthen spatial skills through well-planned training. Cutmore et al. [10] discovered that virtual environments, including 2D projection, can strengthen spatial abilities. Contrarily, the value of AR as a training aid for spatial abilities was not clearly demonstrated in a study by Dünser et al. [13]. The authors [13] claim that AR could be a useful tool for spatial ability training; however, male and female participants had highly distinct training outcomes discovered in their study. Therefore, the creating tools that benefit all users equally, considering individual variations, should be a priority for prospective studies in the near future.

In the present study, we investigated the question: do AR solutions have the potential to significantly improve performance in visuospatial tasks for females and males and overcome inter-sex differences? Although we have gathered a considerable quantity of data, this research focuses on inter-sex differences in perceptions of AR solutions, and their effect on user reaction time and accuracy during the experiment.

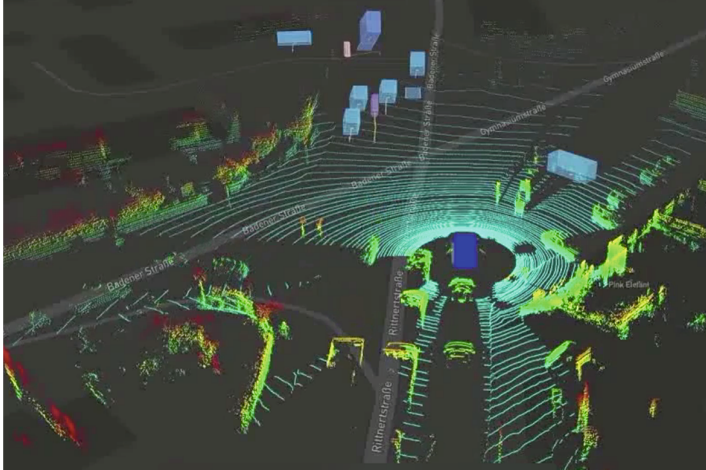
## 3 Methods

### 3.1 Experimental Design

Twenty-two (22) participants (12 women, 10 men, age  $\bar{x} = 29.2$ ) partook in the experiment. Each subject was instructed to detect the appearance of a bicycle and a pedestrian, and subsequently, press a specific button on a keyboard. One session consisted of twenty-two (22) video files lasting less than one minute each. During the study, trials were randomised. Before the experiment, participants performed sample tasks under varying conditions to familiarise them with the required tasks. There was no information on the number of objects and types of visual conditions provided. The results from this study were not anonymous; however, no user names were recorded. Participants were allocated code names (used on all study notes and papers) for data collection to protect participant privacy.

Three experimental conditions were programmed: (1) two targets; (2) one target; (3) only distractors. Trials contained a varying number of targets. With AR features, each moving object was instantly indicated with the appropriate colour box when it occurred on the screen (Fig. 1). The trajectory line was added to anticipate the motion. The colour box disappeared as the target moved behind the vehicle and changed the contrast – becoming bigger compared to distractors farther away, depending on the distance to the car.

In our study, the independent variables were sex and the effective set size within the range of 1 and 17. «Effective set size» is the number of objects that are treated as candidate targets in a specific scene [37]. The difficulty of the task rose concurrently with the increased effective set size.



**Fig. 1.** Example of a 3D visualised moving vehicle and various types of distractors.

## 3.2 Data Analysis

The resulting file included the time stamp and information about the button pressed. Each result was accessed based on reaction time and accuracy. The result was only marked as correct if two requirements were met: (1) the right button was pressed; and (2) the object was detected during its presence on the screen plus an additional 300 ms to allow for decision-making and button pressing. The reaction was only analysed for the correct answers. When calculating the reaction time, we deducted the time of the object's appearance on the screen from the time of the button presses. The final set of analyses included 484 trials (22 trials  $\times$  22 participants).

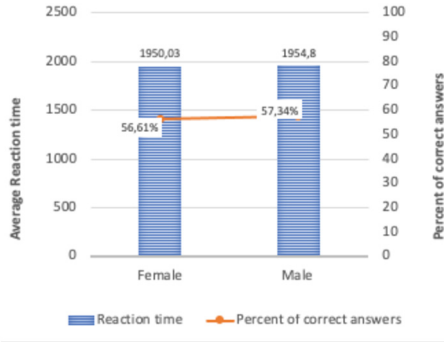
## 4 Results

### 4.1 Percentage of Correct Answers and Reaction Time

We predicted a sex difference in favour of males with a higher accuracy in the completion of the task. In contrast, the average percentage of correct answers was very similar for both sexes, ranging from 56,61% for females to 57,34% for males (Fig. 2).

Further analysis revealed that average reaction time of females (1950,03 ms) closely resembled that of the reaction males (1954,8 ms) (Fig. 2). There was no evident correlation within different age groups (min age = 20, max age = 50) in the percentage of





**Fig. 2.** Sex - Average reaction time and Percent of correct answers

**Table 1.** Indicators depending on sex.

Indicator	Female Median Q1–Q3	Male Median Q1–Q3	p-value
Average reaction time	1661.2 1349.4–2012.8	1777.9 1310.5–2262.5	0.945
Percentage of correct answers	0.7 0.4–0.8	0.7 0.4–0.8	1,000

correct answers (ranging from 60% to 58%). Though, there was some variation in the reaction time (from 1800 ms to 2050 ms) with the increasing age of participants.

We performed a quantitative analysis of the results to assess the relevance of differences between groups using the Mann-Whitney U-test. The variables were presented as medians and quartiles (Q1–Q3). The joint variability of the response variables was evaluated using the Kendall Tau rank correlation coefficient. The critical level of statistical relevance was determined for  $p < 0.1$ . We performed the comparison of indicators “average reaction time” and “percentage of correct answers” depending on sex. The comparison results are presented in Table 1.

**4.2 Effective Set Size and Task Performance**

We analysed the interconnection between the indicators “effective set size”, “average reaction time” and “percentage of correct answers” depending on sex. The correlation results are presented in Table 2 and Figs. 3, 4. The indicators “average reaction time” and “percentage of correct answers” showed a slight negative correlation with “effective set size” ( $p > 0.1$ ). On the other hand, the coefficients of the “male” group skewed slightly more negatively than the “female” group.

According to Fig. 4, the linear trend of average reaction time is almost unchanged with the shift of the effective set size during the experiment, keeping the trend line at approximately 2010 ms. Contrary to our expectations, there is a steady decline in the

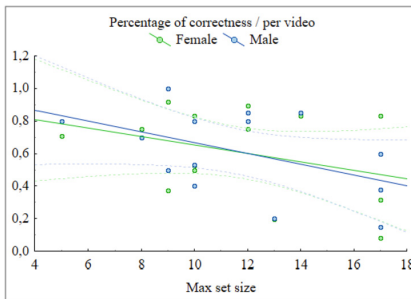
**Table 2.** Correlation of results depending on sex.

Indicator	Female Kendall tau	Female p-value	Male Kendall tau	Male p-value
Average reaction time	-0.150	0.456	-0.242	0.229
Percentage of correct answers	-0.153	0.446	-0.271	0.178

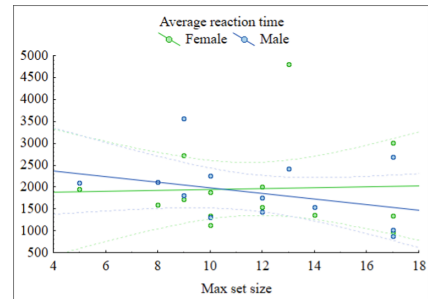
average reaction time for males (starting at 2300 ms to 1540 ms) with increasing task difficulty. Whereas for females, there is a continual growth in the reaction time (starting at 1950 to 2200 ms) correlating with the larger effective set size. Nonetheless, the average reaction time for females (1950,03 ms) was similar to that of the males (1954,8 ms).

## 5 Discussion

Spatial skills include many definite abilities, among which are locating points in space, identifying the orientation of objects, determining location in depth, assessing geometric relations and processing motion [9]. Early studies [25, 29] concluded an acknowledged inter-sex gap in visuospatial ability where males outperform females in spatial tasks; however, this conclusion is much more complex. Voyer et al. [34] showed that some inter-test differences exist, depending on the test used. Males mostly outperformed females in mental rotation tests, while for spatial perception, this advantage diminished [34, 35] assessed that these inter-sex differences in spatial abilities are due to distinctions in visual-spatial working memory, which first appeared in the 13–17-year age group. However, these distinctions varied in magnitude depending on the task effects [34].



**Fig. 3.** “Max set size” and “Percentage of correctness/per video”. Females/Males.



**Fig. 4.** “Max set size” and “Average reaction time”. Females/Males.

For this study, we employed an alternative research design with spatial AR [3], where augmented digital material overlays the target object on the screen. We used this AR-enriched task – which involves object-identification processing and spatial processing – to research the design criteria for AR solutions that can reduce participants’ cognitive load and reaction time and increase their performance.

In their studies on eye tracking and computational models, Carpenter and Just [4, 5, 19, 20] assessed that there is an object-identification process in mental rotation tasks aside

from the rotation and mental transformation of objects. The complex figure is analysed based on its segments (which need to be identified first). Thus, the task involves extensive object identification and comparing objects and their parts [4]. Brain studies confirm that mental rotation tasks connected with the object and object-part identification evoke ITG and MOG areas [18].

In our experiment, we instructed participants to identify moving objects with the help of AR overlaid features. Based on previous studies, we expected a sex-specific difference in accuracy favouring males during the experiment. In contrast, our results showed that accuracy was quite similar for both sexes at a rate of 56,61% for females and 57,34% for males. Further analysis revealed that the average reaction times were also close (1950,03 ms for females and 1954,8 ms for males). Interestingly, the linear trend of the reaction time was mostly unchanged in correlation to increased task difficulty, keeping the average value at 2010 ms. However, a detailed analysis highlighted the inter-sex differences related to task difficulty. There was a slight continual growth in the reaction time of females; in contrast, we observed a steady decline in the reaction time of males. In other words, AR features in our experiment ameliorated inter-sex differences in accuracy and reaction time in visuospatial tasks. However, there might still be differences in cognitive load on a group level as the trend of the reaction time differs significantly. Our results contradict those observed by [2, 36] – where males outperformed females in AR-enriched spatial visualisation and orientation tasks in reaction time, accuracy and navigation tasks. A possible reason for the similar results of females and males in our experiment could be the research design used – AR enrichment is presented directly in the visual field of users. The positive effect of such visual cues is timely and crucial information required to perform the task presented in the participants' field of view [3]. According to Mayer and Fiorella [23], such visual cues guide users' attention to the most relevant data resulting in the decrease of extraneous cognitive load and working memory release. See-through novel AR systems tested by Tsai and Huang [32] and Ro et al. [27] reported lower cognitive load and higher performance compared to traditional systems. Unfortunately, these studies on AR-enriched applications and training and their effects on cognitive load do not assess sex differences.

Therefore, we encourage further studies on AR and cognitive load with inter-sex differences as an additional factor to analyse how computational demand changes during visuospatial tasks for females and males.

## 6 Conclusion

This study reported that AR solutions could ameliorate inter-sex differences in the performance and reaction time in visuospatial tasks involving object-identification and spatial processing. Accuracy decreased for both genders corresponding to increased effective set size and task difficulty. Surprisingly, males showed a steady decrease in reaction time as the complexity of the task increased. Further studies should be considered to assess participants' cognitive load by applying eye-tracking technology.

## References

1. Abramov, I., Gordon, J., Feldman, O., Chavarga, A.: Sex & vision I: spatio-temporal resolution. *Biol. Sex Dif.* **3**, 20 (2012). <https://doi.org/10.1186/2042-6410-3-20>
2. Ahmad, A.M., Goldiez, B.F., Hancock, P.A.: Gender differences in navigation and wayfinding using mobile augmented reality. *Proc. Hum. Factors Ergon. Soc. Ann. Meeting* **49**, 1868–1872 (2005). <https://doi.org/10.1177/154193120504902111>
3. Buchner, J., Buntins, K., Kerres, M.: The impact of augmented reality on cognitive load and performance: a systematic review. *Comput. Assisted Learn.* **38**, 285–303 (2022). <https://doi.org/10.1111/jcal.12617>
4. Carpenter, P.A., Just, M.A., Keller, T.A., Eddy, W., Thulborn, K.: Graded functional activation in the visuospatial system with the amount of task demand. *J. Cogn. Neurosci.* **11**, 9–24 (1999). <https://doi.org/10.1162/089892999563210>
5. Carpenter, P.A., Just, M.A.: Linguistic influences on picture scanning. In: Monty, R., Senders, J. (eds.) *Eye Movements and Psychological Processes*, pp. 459–472. Erlbaum, Hillsdale (1976)
6. Casey, M.B.: Understanding individual differences in spatial ability within females: a nature/nurture interactionist framework. *Dev. Rev.* **16**, 241–260 (1996). <https://doi.org/10.1006/drev.1996.0009>
7. Casey, M.B., Nuttall, R.L., Pezaris, E.: Spatial-mechanical reasoning skills versus mathematics self-confidence as mediators of gender differences on mathematics subtests using cross-national gender-based items. *J. Res. Math. Educ.* **32**, 28 (2001). <https://doi.org/10.2307/749620>
8. Cherney, I.D.: Mom, let me play more computer games: they improve my mental rotation skills. *Sex Roles* **59**, 776–786 (2008). <https://doi.org/10.1007/s11199-008-9498-z>
9. Colby, C.L.: Spatial cognition. In: Squire, L.R. (ed.) *Encyclopedia of Neuroscience*, pp. 165–171. Academic Press (2009). <https://doi.org/10.1016/B978-008045046-9.01120-7>
10. Cutmore, T.R.H., Hine, T.J., Maberly, K.J., Langford, N.M., Hawgood, G.: Cognitive and gender factors influencing navigation in a virtual environment. *Int. J. Hum. Comput. Stud.* **53**, 223–249 (2000). <https://doi.org/10.1006/ijhc.2000.0389>
11. Dietrich, T., et al.: Effects of blood estrogen level on cortical activation patterns during cognitive activation as measured by functional MRI. *Neuroimage* **13**, 425–432 (2001). <https://doi.org/10.1006/nimg.2001.0703>
12. Dirin, A., Alamäki, A., Suomala, J.: Gender differences in perceptions of conventional video, virtual reality and augmented reality. *Int. J. Interact. Mob. Technol.* **13**, 93–103 (2019). <https://doi.org/10.3991/ijim.v13i06.10487>
13. Dünser, A., Steinbügl, K., Kaufmann, H., Glück, J.: Virtual and augmented reality as spatial ability training tools. Presented at the CHINZ 2006: Proceedings of the 7th ACM SIGCHI New Zealand Chapter’s International Conference on Computer-Human Interaction: Design Centered HCI (2006)
14. Durlach, N., et al.: Virtual environments and the enhancement of spatial behavior: towards a comprehensive research agenda. *Presence: Teleoperators Virtual Environ.* **9**, 593–615 (2000). <https://doi.org/10.1162/105474600300040402>
15. Eme, P.-E., Marquer, J.: Individual strategies in a spatial task and how they relate to aptitudes. *Eur. J. Psychol. Educ.* **14**, 89–108 (1999). <https://doi.org/10.1007/bf03173113>
16. Goswami, A., Dutta, S.: Gender differences in technology usage—a literature review. *OJBM* **04**, 51–59 (2016). <https://doi.org/10.4236/ojbm.2016.41006>
17. Hou, L., Wang, X.: A study on the benefits of augmented reality in retaining working memory in assembly tasks: a focus on differences in gender. *Autom. Constr.* **32**, 38–45 (2013). <https://doi.org/10.1016/j.autcon.2012.12.007>

18. Jordan, K., Wüstenberg, T., Heinze, H.J., Peters, M., Jäncke, L.: Women and men exhibit different cortical activation patterns during mental rotation tasks. *Neuropsychologia* **40**, 2397–2408 (2002). [https://doi.org/10.1016/s0028-3932\(02\)00076-3](https://doi.org/10.1016/s0028-3932(02)00076-3)
19. Just, M.A., Carpenter, P.A.: Cognitive coordinate systems: accounts of mental rotation and individual differences in spatial ability. *Psychol. Rev.* **92**, 137–172 (1985). <https://doi.org/10.1037/0033-295x.92.2.137>
20. Just, M.A., Carpenter, P.A.: Eye fixations and cognitive processes. *Cogn. Psychol.* **8**, 441–480 (1976). [https://doi.org/10.1016/0010-0285\(76\)90015-3](https://doi.org/10.1016/0010-0285(76)90015-3)
21. Kimura, D.: Are men's and women's brains really different? *Can. Psychol./Psychologie canadienne* **28**, 133–147 (1987). <https://doi.org/10.1037/h0079885>
22. Linn, M.C., Petersen, A.C.: Emergence and characterization of sex differences in spatial ability: a meta-analysis. *Child Dev.* **56**(138–151), 1479–1498 (1985). <https://doi.org/10.2307/1130467>
23. Mayer, R.E., Fiorella, L.: Principles for reducing extraneous processing: coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In: Mayer, R.E. (ed.) *The Cambridge Handbook of Multimedia Learning*, pp. 279–315. Cambridge University Press (2014). <https://doi.org/10.1017/cbo9781139547369.015>
24. Newcombe, N.S., Mathason, L., Terlecki, M.: Maximization of spatial competence: more important than finding the cause of sex differences. In: McGillicuddy-De Lisi, A.V., De Lisi, R. (eds.) *Biology, Society, and Behavior: The Development of Sex Differences in Cognition*, Greenwood, Westport, pp. 183–206 (2002)
25. Newcombe, N.S.: Taking science seriously: straight thinking about spatial sex differences. In: Ceci, S.J., Williams, W.M. (eds.) *Why aren't There More Women in Science: Top Researchers Debate the Evidence*, pp. 69–77. American Psychological Association, Washington DC (2007). <https://doi.org/10.1037/11546-006>
26. Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., Richardson, C.: A redrawn Vandenberg and Kuse mental rotations test - different versions and factors that affect performance. *Brain Cogn.* **28**, 39–58 (1995). <https://doi.org/10.1006/brcg.1995.1032>
27. Ro, H., Byun, J.-H., Park, Y.J., Lee, N.K., Han, T.-D.: AR pointer: advanced ray-casting interface using laser pointer metaphor for object manipulation in 3D augmented reality environment. *Appl. Sci.* **9**, 3078 (2019). <https://doi.org/10.3390/app9153078>
28. Shaqiri, A., et al.: Sex-related differences in vision are heterogeneous. *Sci. Rep.* **8**, 1 (2018). <https://doi.org/10.1038/s41598-018-25298-8>
29. Stancey, H., Turner, M.: Close women, distant men: line bisection reveals sex-dimorphic patterns of visuomotor performance in near and far space. *Br. J. Psychol.* **101**, 293–309 (2010). <https://doi.org/10.1348/000712609x463679>
30. Tagaris, G.A., Kim, S.-G., Strupp, J.P., Andersen, P., Uürbil, K., Georgopoulos, A.P.: Quantitative relations between parietal activation and performance in mental rotation. *NeuroReport* **7**, 773–776 (1996). <https://doi.org/10.1097/00001756-199602290-00022>
31. Thomsen, T., Hugdahl, K., Ersland, L., et al.: Functional magnetic resonance imaging (fMRI) study of sex differences in a mental rotation task. *Med. Sci. Monit.* **6**, 1186–1196 (2000)
32. Tsai, C.-H., Huang, J.-Y.: Augmented reality display based on user behavior. *Comput. Stand. Interfaces* **55**, 171–181 (2018). <https://doi.org/10.1016/j.csi.2017.08.003>
33. Unterrainer, J., Wranek, U., Staffen, W., Gruber, T., Ladurner, G.: Lateralized cognitive visuospatial processing: is it primarily gender-related or due to quality of performance? *Neuropsychobiology* **41**, 95–101 (2000). <https://doi.org/10.1159/000026639>
34. Voyer, D., Voyer, S., Bryden, M.P.: Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychol. Bull.* **117**, 250–270 (1995). <https://doi.org/10.1037/0033-2909.117.2.250>

35. Voyer, D., Voyer, S.D., Saint-Aubin, J.: Sex differences in visual-spatial working memory: a meta-analysis. *Psychon. Bull. Rev.* **24**, 307–334 (2017). <https://doi.org/10.3758/s13423-016-1085-7>
36. Waller, D., Hunt, E., Knapp, D.: The transfer of spatial knowledge in virtual environment training. *Presence* **7**, 129–143 (1998). <https://doi.org/10.1162/105474698565631>
37. Wolfe, J.M., Horowitz, T.S.: Five factors that guide attention in visual search. *Nat. Hum. Behav.* **1** (2017). <https://doi.org/10.1038/s41562-017-0058>
38. Yan, Z., Shan, Y., Li, Y., Yin, K., Li, X.: Gender differences of cognitive loads in augmented reality-based warehouse, pp. 500–501 (2021). <https://doi.org/10.1109/vrw52623.2021.00132>



# Parkinson's Disease Assessment from Speech Data Using Recurrence Plot

Arsya Mohamed Ali<sup>1</sup>, G. Jyothish Lal<sup>1</sup>(✉), V. Sowmya<sup>1</sup>,  
and E. A. Gopalakrishnan<sup>2</sup>

<sup>1</sup> Center for Computational Engineering and Networking (CEN),  
Amrita Vishwa Vidyapeetham, Coimbatore, India  
cb.en.p2dsc21006@cb.students.amrita.edu,  
{g.jyothishlal,v.sowmya}@cb.amrita.edu

<sup>2</sup> Amrita School of Computing, Amrita Vishwa Vidyapeetham, Bangalore, India  
ea.gopalakrishnan@blr.amrita.edu

**Abstract.** Brain diseases, which encompass a wide range of conditions and illnesses brought on by stroke, Alzheimer's, Parkinson's, traumatic brain injury, and many other conditions, afflict 1 in 6 individuals worldwide. One such disorder that gradually destroys nerve cells in the mid-brain is Parkinson's disease. It is a neurodegenerative condition that impairs motor abilities. Additionally, it has an impact on the muscles involved in speech production, leading to hypokinetic dysarthria, a set of motor speech disorders that includes dysphonia, bradylalia, and poor articulation accuracy. In this context, this work proposes the application of recurrence plots to map the speech signals onto images, which will be further fed into a Convolutional Neural Network for the automatic classification of PD from healthy controls. The proposed approach is assessed on the Italian Voice and Speech Data, containing Diadochokinetic (DDK) recordings of /pa/ & /ta/ audio recordings. The experimental results of the full audio file approach produced an average testing accuracy of 83%. Also, the model based on the frame-based approach performed well on the test set, resulting in an average test accuracy of 91% for both /pa/ and /ta/ recordings.

**Keywords:** Complex System · DDK Analysis · Hypokinetic Dysarthria · Convolutional Neural Network · Parkinson's Disease · Recurrence Plot

## 1 Introduction

Approximately 1 million people are currently affected with Parkinson's disease, a chronic, degenerative neurological condition. By 2030, it is anticipated that this number will reach 1.2 million [3]. After Alzheimer's disease, Parkinson's disease (PD) is the most common neurodegenerative condition. PD affects over 10 million people globally and is more common in older people. Only 4% of cases

are diagnosed before the age of 50, and it affects men 1.5 times more often than women [3].

World Health Organisation (WHO) states that this degenerative condition of the brain is associated with motor symptoms and a wide variety of non-motor complications. Speech and mobility difficulties, as well as limitations in many facets of daily living, are caused by motor disorders such as dyskinesias (uncontrollable movements) and dystonias (painful uncontrollable muscular contractions) [1, 2].

As PD impacts non-motor functioning, it is known that around 90% of patients experience various speech deficits. This includes reduced loudness, a mono-pitch and mono-loudness kind of speech, breathy voice, and imprecise articulation, among others. Hypokinetic dysarthria is a collective term for all of these signs and symptoms [11]. As speech signal is a bio-marker to detect PD, this work primarily aims to classify a patient as PD or Healthy Control (HC) by training 2DCNN (2D-Convolutional Neural Network) with the recurrence plots extracted from speech data corresponding to PD and HC individuals.

In this research, a complex system approach utilizing recurrence plots (RP) is employed to analyze speech signals for the purpose of categorizing patients as either healthy control (HC) or suffering from Parkinson's disease (PD). The purpose is to characterize the variation in the dynamics of the speech system during the repeated pronunciation of /pa/, and /ta/ by PD patients and HC. DDK (Diadochokinetic) analysis is a measure to assess the ability to make rapidly alternating speech movements by SLPs (Speech Language Pathologists). Here in this research work, the audio files containing the DDK exercise of the syllables /pa/ and /ta/ are considered.

## 2 Literature Review

The automatic classification of speakers with PD and HC has been the subject of several investigations. Most of the research works discussed in this section used speech recordings of sustained vowels, lone words, short phrases, and spontaneous speech as their primary data. To date, most of the works that have been reported for classifying a patient as having PD have used the acoustic and cepstral features of the speech recordings to train the models. There have been studies that used network parameters and recurrence plots to train ML and DL models, but they were not done on Parkinson's speech recordings.

The paper [7] presents a manual approach for assessing voiceless plosive VOT and examines its effectiveness in identifying specific articulation issues in PD patients. The approach includes measuring VOT values, consonant length, closure phase duration, and peak spectral energy during the burst phase. The study used the PC-GITA Database's DDK /pa-ta-ka/ recordings that is the rapid repetition of the syllables 'pa', 'ta', and 'ka' in a rhythmic pattern [5]. The best accuracy of 77% was achieved using SVM-RBF trained with features from the consonant /pa/ and the three consonants.

The study [8] compares various models to assess phonation symptoms in PD patients' using the PC-GITA database. The models include features from



zero frequency filtered (ZFF) signals, perturbation, and glottal features. Raw waveform CNNs are also applied to extract features from speech waveforms, glottal wave reconstructions, and ZFF signals. The results showed that SVM with a Gaussian kernel model accurately distinguished low and high phonatory impairments, achieving 76–84% accuracy.

The research work [9] proposes pre-trained GRUs to automatically detect Voice Onset Time (VOT) and analyze how motor speech disorders affect speech production. The recordings include the DDK task utterance /pa-ta-ka/ from the PC-GITA Database. Acoustic analysis is performed by extracting distinct temporal and spectral elements, with 42 features considered (18 temporal, 24 spectral). Manual VOTs are trained using the BiGRU. The trained model can detect VOT with F1-score values of 0.66 for /pa/, 0.75 for /ta/, and 0.78 for /ka/ when compared to manual VOT labels.

The authors of [6] research use the framework of complex networks to examine the non-linear behavior in the dynamics of voiced and unvoiced speech production. The recurrence network approach transforms the time series corresponding to a speech utterance into a complex network or graph. This is a novel approach to speech signal processing that aims to distinguish between voiced and unvoiced speech by comparing three network parameters: CPL, CC, and degree distribution.

The Recurrence Plot (RP) introduced by Eckmann et al. in their work [12] states that RP enables visualization of repetitive events of higher dimensions through projections onto a 2- or 3-dimensional representation. Luis et al. in their work [13] used signals recorded during specific drawings by PD patients and HC to train different CNN models after converting them into recurrence plot images, achieving an average accuracy of over 87% and demonstrating the potential of recurrence plots for training DL models in identifying Parkinson's patients.

Some studies have used non-linear behavior of data to build recurrence or complex networks. Godavarthi et al. utilized recurrence networks to explore dynamical transitions in a turbulent combustor and found significant differences in network structures between time series of acoustic pressure and white noise [10]. In another study, researchers used a complex network approach to study plant growth by defining pixels of images as network nodes. The constructed complex network was used to examine the spectral patterns and characteristics of images, and the network was able to distinguish between healthy, unhealthy, and moderate plants [16].

This research aims to address the gap in the existing literature by implementing a recurrence plot (RP)-based approach on speech to detect PD. Previous studies did not explore the possibility of using RP of speech utterances of PD patients as an image to build a classification model. The primary question that this research attempts to answer is “Will the 2DCNN trained with RP be able to distinguish between speech recordings of PD patients and HC?”. This research aims to use RP analysis to study dynamical transitions in speech recordings of PD and HC patients and focuses on building a 2DCNN model for classification.

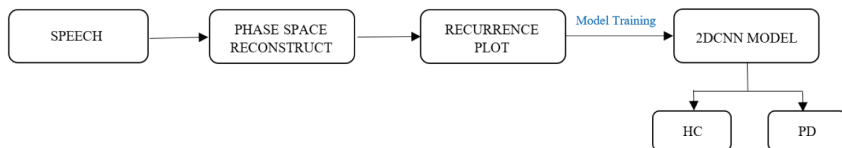
### 3 Materials and Methodology

#### 3.1 Dataset Description

The Italian Voice and Speech Dataset used in this work is cited in the paper titled “Assessment of Speech Intelligibility in Parkinson’s Disease Using a Speech-To-Text System” [4]. This dataset includes the audio recordings of 65 individuals, including 25 people with PD, 22 old healthy controls, and 15 young, healthy controls. For this work, audio files of the subject reading the syllable “pa” for five seconds, followed by a 20-second gap, and the word “ta” for five seconds are used. It should be noted that there is a one-minute pause between executions. In this set of audio files, only 50 subjects’ audio files are considered, which include 22 elderly healthy controls and 25 PD patients. The recordings were downsampled to 16 kHz to reduce the computational complexity.

#### 3.2 Methodology

To get an RP for each audio file, the first step is to reconstruct the phase space of the speech data. By considering time series by the time-delay embeddings, the phase space plot trajectory can be reconstructed [14]. Figure 1 shows the workflow of the proposed method.



**Fig. 1.** Proposed Workflow

The reason for creating RP is that many dynamical systems exhibit recurrence, which is a fundamental characteristic in which the system strives to return to a previous state. RP is a tool that can be used to show the recurrence of states of a system [14]. Consequently, the phase space must be recreated using the time-series data. All possible states of a system are represented in the phase space, with each conceivable state corresponding to a distinct point.

By selecting an appropriate time delay value  $\tau$  and embedding dimension  $d$  for a time-series  $[x_1, x_2, x_3, \dots, x_N]$  of length  $N$ , we can embed the time series into higher-dimensional phase space [6]. This allows us to describe the created delay-coordinate vectors as

$$x_i = [x(i), x(i + 1), x(i + 2), \dots, x(i + (d - 1)\tau)] \quad (1)$$

where  $i = 1, 2, \dots, N - (d-1)\tau$ . The embedding parameters  $\tau$  and  $d$  are suitably defined for accurate reconstruction. Calculating the first minimum of the average

mutual information (AMI) yields the ideal time delay  $\tau$ . The least embedding dimension  $d$  is calculated using the false closest neighbors (FNN) method [17]. Mathematically, RP can be expressed as

$$R_{i,j} = \theta(\epsilon_i - \|\bar{x}_i - \bar{x}_j\|), \quad \bar{x}_i \in R^m, \quad i, j = 1, \dots, N, \quad (2)$$

where  $x_i, x_j$  are the states at  $i$  and  $j$  respectively,  $\theta$  is the Heaviside function whose value will be 1 for positive input and zero for others, and  $\epsilon$  is the threshold distance. Each point in phase space is regarded as a node, and their connection is based on the recurrence threshold  $\epsilon$ . As a result, a square matrix is created, with each entry set to 1 if the pair-wise distance is less than the threshold distance  $\epsilon$  and zero otherwise. In other words, if the phase space distance between two nodes is less than  $\epsilon$ , they are connected.

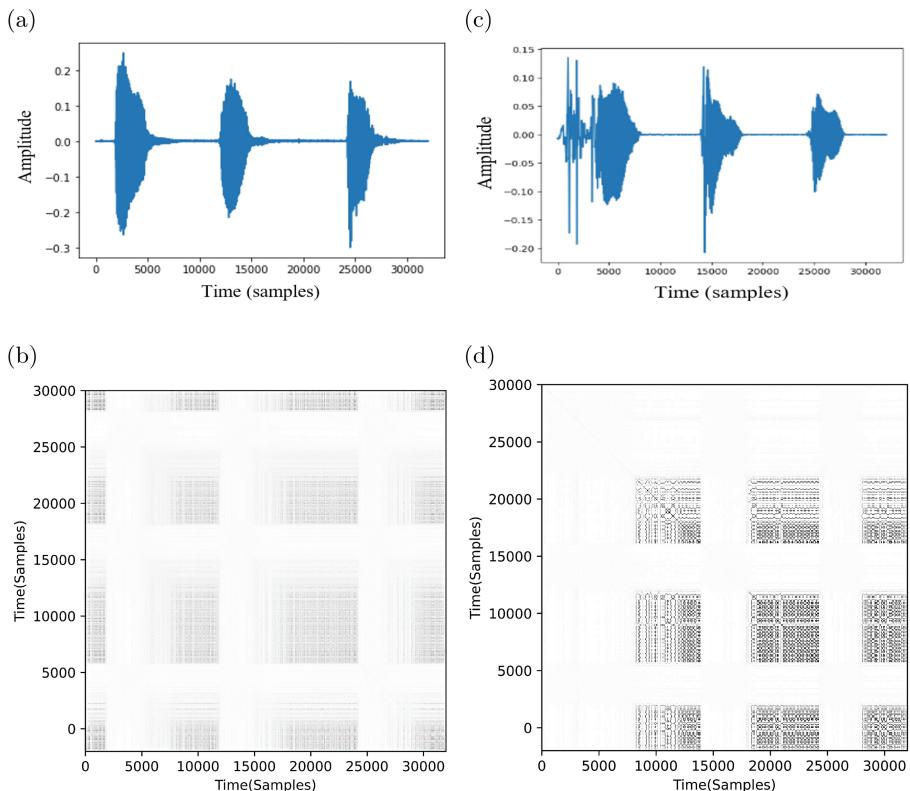
The creation of RP from phase space plots is automated using the Pyunicorn library in Python [15], with the recurrence rate set at 0.05 [6]. This study explores two approaches: one is to consider full-length audio files, and the other is frame-based analysis.

**Full-Length Audio File Approach:** The approach involves creating RPs from full-length audio files, using DDK recordings of /pa/ and /ta/ from the Italian dataset. 50 RPs are created (22 from healthy controls and 28 from PD patients). Figure 2 shows a sample visualization of the RPs for both HC and PD. Visual observation reveals that there are more disruptions (white bands) in the RP corresponding to PD patient when compared to that of the HC. This indicates that some of the states in the system corresponding to PD patient is far from normal, and it is possible that transitions have taken place [14].

**Frame-Based Analysis:** By this approach, the dynamics in each frame of the recording can be captured, which might show a significant difference in the RP of HC and PD. We have chosen a frame length of 50 ms (which corresponds to 800 samples for  $F_s = 16$  kHz) and opted for a half overlap of 25 ms (which corresponds to 400 samples for the same  $F_s$ ) [6]. The duration is restricted to 2s to make the frame length equal across the dataset.

## 4 Model Architecture

As discussed in Sect. 3, two approaches have been experimented with in this work, that is, the full audio file approach and the frame-based analysis. In order to develop a 2DCNN classification model using recurrence plot images, the following architectures have been used. Figure 3(a) is the architecture used for the full audio file approach. This architecture consists of an input layer followed by three sets of convolution and max pooling layers and each convolutional layer consists of about 8, 16, and 32 neurons, respectively, with relu activation function, L1 kernel regulariser, and Glorot uniform initializer. Following this, it has

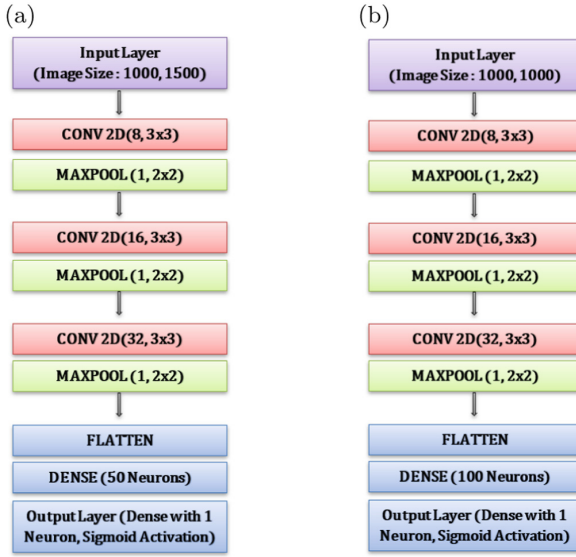


**Fig. 2.** Visualization of RP's. (a) Speech signal corresponding to HC /papapa/ and (b) its corresponding RP, (c), (d) similar plots for PD patient /papapa/

a flattened layer, followed by a dense layer of 50 neurons and an output layer with a sigmoid activation function.

The frame-based approach in this research uses the architecture shown in Fig. 3(b), which includes an input layer, three sets of convolution and max pooling layers, a flattened layer, a dense layer with 100 neurons, and an output layer with sigmoid activation. The convolutional layers have 8, 16, and 32 neurons with relu activation, L1 kernel regularizer, and Glorot uniform initializer.

The lambda layer in both architectures rescales the input image to a range of 0 to 1, and the default kernel initializer is used to initialize weights for the neural network. Binary cross entropy is used as the loss function, and the Adam optimizer is used.



**Fig. 3.** Model Architecture for (a) Full Audio File Approach and (b) Frame - Based Analysis

## 5 Model Training and Results

**Table 1.** Number of HC and PD Patients in Train, Test, and Validation Sets

Class	Train Set	Test Set	Validation Set
HC	16	3	3
PD	21	3	4

The model training for both approaches was carried out considering different train-test splits like 75:25 and 80:20 and also with and without dropouts, regularizer, and initializer. The model with the best results has been recorded and shown in this work.

### 5.1 Full Audio File Approach

For this approach, the architecture as in Fig. 3(a) was considered, and the model was trained for 20 epochs for both /pa/ and /ta/ DDK recordings. The train-test split was taken as 75:25, in which the test set is again split into half for the validation set. Table 1 shows the number of HC and PD patients in each set. The evaluation results are shown in Table 2.

**Table 2.** Model Evaluation Results

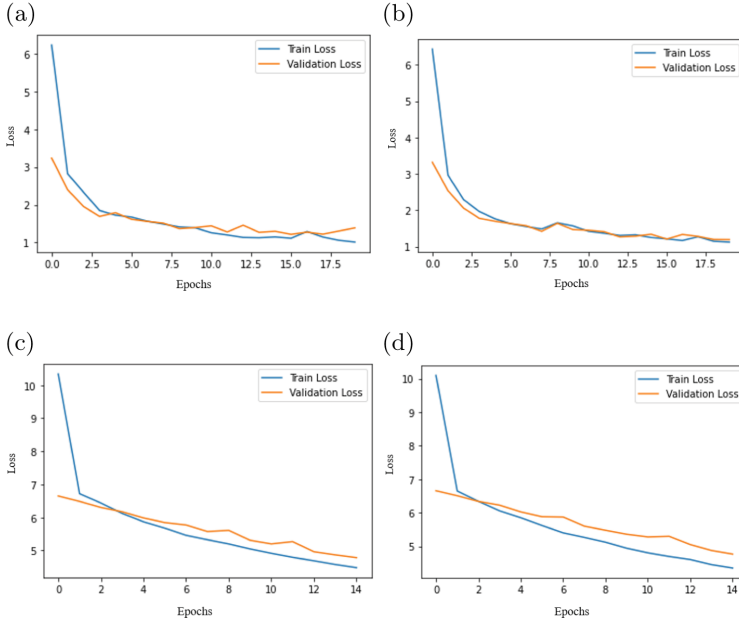
Approach	Dataset	Accuracy(in %)			F1 Score
		Train	Test	Validation	
Full Audio	DDK /pa/	89.19	83	71.4	0.80
	DDK /ta/	86	83.3	100	0.85
Frame-Based	DDK /pa/	100	83	85.7	0.85
	DDK /ta/	100	100	85.7	1

## 5.2 Frame-Based Analysis

In this approach, the model was trained for 15 epochs using Fig. 3(b) architecture for both /pa/ and /ta/ DDK recordings. The frames of a patient were grouped together without shuffling, resulting in patient-wise training with a shape of (50, 1000, 1000, 79). The shape means that there are 50 patients, each patient having 79 frames, and each frame is of size  $1000 \times 1000$ . The train-test split was taken as 75:25 and as frames are passed patient-wise, the split is the same as in Table 1. The evaluation results are shown in Table 2, where each model’s train accuracy, test accuracy, and validation accuracy are listed.

As the test and validation samples are fewer, even if one sample is misclassified, the accuracy will drop to 0.8. Hence, taking accuracy as the evaluation metric can be misleading, so loss has been taken into consideration. The train and validation losses for each scenario are shown in Fig. 4. The loss function measures classification performance, and decreasing train and validation losses with increasing epochs show the model’s effectiveness in the classification task on both training and validation sets.

The F1 score measures a model’s ability to accurately identify positive cases while maintaining precision in its predictions. From Table 2, by considering the test accuracy and F1 score of Italian /pa/ and /ta/ DDK recordings, the model trained with RPs of the full audio file classified PD from HC with 83% test accuracy and f1 scores of 0.8 and 0.85 for /papapa/ and /tatata/ test sets, respectively, with only one misclassification. The frame-based analysis model also performed well on the test set as the /papapa/ model had one misclassification, and the /tatata/ model had no misclassification, contributing to the test accuracy of 83% and 100% and F1 score of 0.85 and 1, respectively.



**Fig. 4.** Model Loss Plots. Full Audio File Approach loss plots for (a) Italian /papapa/ and (b) Italian /tatata/, (c, d) similar plots for Frame-Based Analysis

## 6 Conclusion

The proposed work explored the usefulness of a complex system-based approach, namely, the recurrence plots for the classification of PD and HC individuals from speech data. We explored two approaches to create recurrence plots, and the model trained using both methods showed promising results. Still, the Full audio approach was better in terms of time and space complexity for RP generation and training. The frame-based approach achieved 83% accuracy on the /pa/ test set with one misclassification, and 100% accuracy on the /ta/ set with no misclassifications in fewer epochs than the full audio model. Taking into account time, space complexity, and training results, the full audio model achieved 83% accuracy in classifying HC and PD patients. As per the objective of this work, 2DCNN trained with RP images of HC and PD speech signals showed promising results which in the future, can be experimented for various research objectives. Here in this work, a frame length of 50 ms is used in reference to the paper [6]; instead, different frame lengths can be tried, and results can be recorded. Models can be used to evaluate other language recordings, and the proposed approach can be tested on DDK recordings of /ka/, /pataka/, /petaka/, and /pakata/.

## References





1. Parkinson's Disease Risk Factors and Causes—Johns Hopkins Medicine. <https://www.hopkinsmedicine.org>. Accessed 2 Dec 2022
2. Parkinson disease—World Health Organisation. <https://www.who.int>. Accessed 2 Dec 2022
3. Statistics—Parkinson's Foundation. <https://www.parkinson.org>. Accessed 2 Dec 2022
4. Dimauro, G., De Nicola, V., Bevilacqua, V., Caivano, D., Girardi, F.: Assessment of speech intelligibility in Parkinson's disease using a speech-to-text system. *IEEE Access* **5**, 22199–22208 (2017)
5. Orozco, J.R., Arias-Londoño, J.D., Vargas-Bonilla, J., González-Rátiva, M., Noeth, E.: New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease. In: Proceedings of the 9th Language Resources and Evaluation Conference (LREC), pp. 342–347 (2014)
6. Jyothish Lal, G., Gopalakrishnan, E.A., Govind, D.: Glottal activity: a recurrence network approach for characterization and detection of dynamical transitions during human speech production. *Circuits Syst. Signal Process.* **41**, 6975–6998 (2022)
7. Argüello-Vélez, P., Arias-Vergara, T., González-Rátiva, M.C., Orozco-Arroyave, J.R., Nöth, E., Schuster, M.E.: Acoustic characteristics of VOT in plosive consonants produced by Parkinson's patients. In: Sojka, P., Kopeček, I., Pala, K., Horák, A. (eds.) *Text, Speech, and Dialogue. LNCS*, vol. 12284, pp. 303–313. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-58323-1\\_33](https://doi.org/10.1007/978-3-030-58323-1_33)
8. Vásquez-Correa, J.C., Fritsch, J., Orozco-Arroyave, J.R., Nöth, E., Magimai-Doss, M.: On modeling glottal source information for phonation assessment in Parkinson's disease. In: Proceedings of the Interspeech, pp. 26–30 (2021). <https://doi.org/10.21437/Interspeech.2021-1084>
9. Arias-Vergara, T., et al.: Automatic detection of Voice Onset Time in voiceless plosives using gated recurrent units. *Digital Signal Process.* **104**, 102779 (2020). <https://doi.org/10.1016/j.dsp.2020.102779>. ISSN 1051-2004
10. Godavarthi, V., Unni, V.R., Gopalakrishnan, E.A., Sujith, R.I.: Recurrence networks to study dynamical transitions in a turbulent combustor. *Chaos* **27**(6), 063113 (2017). <https://doi.org/10.1063/1.4985275>
11. Orozco-Arroyave, J.R., et al.: Voiced/unvoiced transitions in speech as a potential bio-marker to detect Parkinson's disease. In: Proceedings of the Interspeech 2015, pp. 95–99 (2015). <https://doi.org/10.21437/Interspeech.2015-34>
12. Eckmann, J.P., Kamphorst, S.O., Ruelle, D.: Recurrence plots of dynamical systems. *Europhys. Lett.* **9**(4), 973–977 (1987)
13. Afonso, L.C.S., et al.: A recurrence plot-based approach for Parkinson's disease identification. *Future Generation Comput. Syst.* **94**, 282–292 (2019). <https://doi.org/10.1016/j.future.2018.11.054>. ISSN 0167-739X
14. Marwan, N., Romano, M.C., Thiel, M., Kurths, J.: Recurrence plots for the analysis of complex systems. *Phys. Rep.* **438**(5–6), 237–329 (2007). <http://www.recurrence-plot.tk>. Accessed 20 Apr 2023
15. pyUnicorn - UNified COMplex Network and Recurrence aNalysis toolbox. <https://tocsy.pik-potsdam.de>. Accessed 10 Aug 2022



16. Sajith, V.V.V., Gopalakrishnan, E.A., Sowmya, V., Soman, K.P.: A complex network approach for plant growth analysis using images. In: 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, pp. 0249–0253 (2019). <https://doi.org/10.1109/ICCSP.2019.8698021>
17. Jyothish Lal, G., Gopalakrishnan, E.A., Govind, D.: Glottal activity detection from the speech signal using multifractal analysis. *Circuits Syst. Signal Process.* **39**, 2118–2150 (2020)



# COVID-19 and Behavioral Analytics: Deep Learning-Based Work-From-Home Sensing from Reddit Comments

M. M. Enes Yurtsever<sup>1</sup> , Ekin Ekinci<sup>1</sup> , and Süleyman Eken<sup>2</sup>  

<sup>1</sup> Kocaeli University, 41001 Izmit, Kocaeli, Turkey  
enes.yurtsever@kocaeli.edu.tr, ekinekinci@subu.edu.tr

<sup>2</sup> Sakarya University of Applied Sciences, 54050 Serdivan, Sakarya, Turkey  
suleyman.eken@kocaeli.edu.tr

**Abstract.** For the foreseeable future, managing the work from home (WFH) workplace will continue to be a challenge for enterprises. An essential stage in developing and improving a plan is utilizing topic modeling and polarity sentiment analysis. This strategy can assist executives in sustaining productivity, customizing employee settings and experiences, and sustaining successful teamwork. The objective of this research is to utilize Reddit's subreddits and behavioral analytics to detect work-from-home trends in the context of COVID-19, employing advanced computational methods. On textual data collected from 7 communities focusing on WFH-related discussion on Reddit from December 1, 2020, to August 31, 2022, we use sentiment analysis and latent topic modeling. Polarity analysis revealed that communities had a more favorable attitude towards WFH than a negative one. Community members were found to be worried about a range of issues associated to remote employment, according to topic modeling.

**Keywords:** COVID-19 · Deep learning · Digital nomad · Sentiment analysis · Topic modeling

## 1 Introduction

Living things continually sense and assess the environment around them. This applies to both live things and inanimate items. All of it is done with the intention of drawing conclusions and acting—consciously or not—in some way. Conducting daily activities and providing support for other tasks require an understanding of our surroundings and how others behave. There are three basic justifications for wanting to automatically assess behaviors: (i) React-based on what is occurring in the environment, a biological or artificial agent (or a combination of both) may behave. (ii) Understand - It is possible to comprehend other related behaviors and processes and find answers to research problems by analyzing an organism's behavior. (iii) Document and Archive - Last but not least, we might want to record particular actions for later use. It might be for evidence, or perhaps it's unclear how the information can be used right now, but it might be useful in the future [1].

The tracking, gathering, and evaluation of user data and behaviors via monitoring systems is known as user behavior analytics (UBA). In order to reflect the fact that users are merely one type of entity with observable behaviors on contemporary networks, UBA is frequently referred to as user and entity behavior analytics (UEBA). Processes, applications, and network devices are examples of other entities [2]. Also, trend analytics in people's work behavior is one type of UBA. In particular, the main trigger of this trend has been the COVID-19 pandemic. In this paper, we utilize the soft computing methods which yield promising solutions for behavioral analytics.

The SARS-CoV-2 pandemic has had a devastating impact on people, organizations, and governments all across the world [3]. The most significant pandemic of our time is COVID-19, which has produced over 493 million illnesses and nearly 6.2 million fatalities worldwide<sup>1</sup> as of this writing. The COVID-19 virus's global proliferation and ongoing persistence [4] have had significant social, economic, and health repercussions. The virus's unpredictable spread and return in many forms has made it difficult to manage and control it as well as deal with its effects. The need to strike a balance between containment and mitigation efforts and economic and social revivals has made policymakers more reluctant to rely solely on these strategies. Nowadays, the majority of businesses use remote or WFH policies. In order to support flexible working arrangements, remote work is defined as work that employees accomplish at home utilizing information and communication technology. In sophisticated economies, 20 to 25 percent of the workforce might work remotely without suffering from a productivity loss. The tendency of remote work being the new norm will persist beyond the pandemic [5]. This study explores the following research questions:

The subsequent research inquiries are examined in this study:

- RQ1: On Reddit, the most popular social media site, what is the overall feeling or opinion about working from home (WFH)?
- RQ2: According to Reddit comments, what are the primary WFH-related subjects?
- RQ3: What is the polarity-based change in latent topics?
- RQ4: How do latent subjects vary on a monthly basis?
- RQ5: Where do comments tend to appear most frequently?

The contributions of this work to the literature are as follows:

- We use different number of subreddits on Reddit to sense the "work from home" related topics to observe UBA triggered by COVID-19.
- We characterize differences in WFH topics and sentiments.

The remainder of this paper is organized as follows. Section "Related Works" includes three sub-sections: (i) WFH and its challenges and benefits, (ii) Behavioral perspectives from human, organizational, and business, and (iii) Sentiments towards WFH. We give the details of the proposed work in Section "Materials and Methods". Section "Results and Discussion" firstly gives our own created dataset and secondly performance results. Finally, the last section presents the "Conclusion and Future Works" of this paper, states the significance of it, and gives future works.

## 2 Related Works

### 2.1 WFH and Its Challenges and Benefits

WFH has some benefits and challenges. Flexibility is the pervasive advantage of virtual working. It enables workers to adjust their working hours in accordance with their schedule and take care of activities that call for them to be at home [6]. Additionally, workers can modify the lighting, warming, setting, and background noise in their workspace to suit their needs. Who wants to spend hours traveling to and from work? Another significant benefit of WFH is that it minimizes the need for a commute, saving time and money and boosting morale. Because they have more time to devote to looking after their physical or mental health, these time savings can also lead to a better work/life balance [7]. From the standpoint of productivity, virtual workers are claimed to produce more work since meetings are more efficient and there are fewer coworker-related interruptions. As a remote worker, they might also discover that they take fewer sick days because they are less likely to take a day off for a minor sickness. It is simpler to manage their workload and produce results effectively and efficiently when they take fewer vacation days [8].

There are difficulties associated with remote work. To begin with, working from home may be extremely alienating and result in long stretches of time spent alone. In keeping with this, it might be challenging to keep company's culture and employee community alive. There are less opportunities to interact, participate, and develop a sense of camaraderie when there is no way to have a cup of tea with a colleague or check in to see how they are doing. Being away from the office atmosphere also lessens your exposure to the broader mission and values of the organization [9]. Switching off can be challenging when working virtually since the lines between work and play can become hazy. Zapier's research indicates that remote workers are more prone to overwork because it is more difficult to ignore duties after hours without the structure of an office environment. Some people find that working from home increases their risk of procrastination or distraction since a fast task at home takes longer than intended or they become sidetracked by a visitor or young children. Additionally, it limits performance evaluation and, in some cases, raises the risk of being passed over for promotions or career advancement.

### 2.2 Behavioral Perspectives from Human, Organizational, and Business

It is crucial to comprehend and examine the COVID-19's impacts on humans. For this reason, we give the ramifications that this pandemic is having on the workplace in this sub-section so that study strategies pertaining to problems with organizational, business, and human behavior can be developed. Since the COVID-19 outbreak forced the strategy processes to adapt, many firm managers have had to sail through difficult waters with expensive consequences. Supporting partners, clients, and suppliers is a top responsibility right now in order to prevent supply chains in which they are involved from being disrupted. In light of uncertain scenarios, it is important to emphasize the importance of an entrepreneur's vision, skills, and leadership so they can recognize the earliest consequences of difficult circumstances and build survival tactics. From a commercial

standpoint, the urgency and complexity of the new normal were being taken into consideration. In order for a firm to overcome uncertainty, it must adhere to strong health rules that provide all customers and collaborators with trust and security [10].

Many of the biases in human behavior are the consequence of first impressions, while others are ingrained habits that are the result of repetition and associative learning. Due to the COVID-19 pandemic, behavioral economics has assumed a crucial role in putting an end to the pandemic. By altering how choices are given to people when making decisions, it is required to alter their behavior. Behavioral economics examines how psychological, social, and cognitive factors influence people's decisions and incorporates concepts from the fields of psychology [11], neurology, and microeconomics. It is important to encourage acceptable changes in human behavior through the creation of simple tools and messaging [12]. Numerous organizations have the chance to create new strategies and the capabilities to do so, including becoming more digital, streamlining processes, automating workspaces, designing cost structures that are more flexible, responding to decision-making, having the right technological support, and developing secure alternatives for online distribution, delivery, and commerce. Business executives ought to take these concepts into consideration [13]. Whatever the future holds for remote work, it is highly possible that many organizational practices, policies, and ideologies will be based on the digital footprints created during COVID-19. In order for organizational researchers to offer the insights we need to carefully create and manage the organizations of the future, it is imperative that we comprehend, debate, and analyze these second-order implications of remote labor now [14].

### 2.3 Sentiments Toward WFH

A great data source for leveraging understanding of an issue from a public perspective can be tapped into thanks to the massive amount of information created on social media platforms. Additionally, earlier studies recommend gathering and analyzing data posted on social media platforms in order to identify and address concerns in real-time in order to improve quality of life. To increase managers' awareness of real-world events like WFH during the pandemic, acquire larger people's attitudes, and access segments that are less inclined to participate in standard types of data collecting, data mining of the real-time social media posts is necessary. Some researchers focused on sentiments toward WFH using social media data.

Wrycza and Maślankowski [15] assessed remote work phenomenon by analyzing tweets on Twitter. They came to the conclusion that the topic of remote work reached a pandemic peak in March 2020 and that it had increased nearly 15 times over the previous year. The sentiment research also revealed that more than 60% of users approved of remote work. The study validated the idea that remote work will continue to exist after COVID-19. Dubey and Tripathi [16] analyzed 100,000 tweets posted on social media to see how people felt about the WFH concept. Results showed that respondents accepted the WFH idea well. The majority of the tweets' feelings were ones of trust and anticipation, showing that people are accepting of this idea. Natural language processing (NLP) methods were used to analyze 1 million tweets from March 30 to July 5 of 2020 in Zhang et al.'s study [17], which aimed to determine how the general public felt about remote work. The findings revealed generally positive opinions expressed in tweets

about remote employment, with slight lulls over the weekend. The findings of topic modeling indicated recurring themes in tweets, including home offices, cybersecurity, mental health, work-life balance, teamwork, and leadership. Over the course of the 14-week period, these themes only slightly changed. Daneshfar et al. [18] sought to better understand the difficulties associated with WFH in the context of the ongoing COVID-19 pandemic, to look into how the general public felt about this change, and to develop a conceptual model incorporating the relationships between the variables that affect WFH's effectiveness. They employed the netnography approach to gather information from the Twitter platform, and they ran sentiment analysis and directed content analysis on the information using the Python programming language, NLP techniques, and IBM SPSS 26. The bulk of tweets concerning WFH during the pandemic were found to be objective and positive, with technology and cyber security being the most often discussed subjects. Table 1 summarizes the state-of-the-art comparison.

**Table 1.** State-of-the-art comparison.

Work	Social media platform	COVID stage	Methodology	UBA
Wrycza and Maślankowski [14]	Twitter	During	Text mining K-means Naïve Bayes	X
Dubey and Tripathi [15]	Twitter	During	Sentiment analysis	X
Zhang et al. [16]	Twitter	During	Latent topic modelling Sentiment Analysis	X
Daneshfar et al. [17]	Twitter	During	Netnography approach NLP techniques and IBM SPSS 26	X
<b>Ours</b>	Reddit	During and post	Sentiment analysis Latent topic modeling Word embedding	✓

### 3 Materials and Methods

#### 3.1 Data Collection

To sense WFH in the context of COVID-19 and UBA, we collected information from the Reddit information-sharing social media network, which has around 430 million active members at this time. From 7 subreddits (digitalnomad, homeoffice, Remote-Jobs, remotework, WFH, workfromhome, and WorkOnline), we gathered about 14,512 comments. 'Freelance, remote' as search terms brought up the most pertinent responses.

### 3.2 Data Pre-processing

Raw data were subjected to a number of pre-processing operations. First, the Reddit comment body column was changed to a string type, and extra columns, NaN, and NaT rows were removed. Then, using regular expressions, all Reddit comments were made lowercase and non-alphabetic letters, URLs, hyperlinks, emojis [19], special characters, extra new lines, and user references were eliminated. Following that, texts were lemmatized with spaCy and cleaned of stop words with Gensim [20]. After the data was cleaned and organized, we performed sentiment analysis and NMF topic modeling using NLP tools in Python.

### 3.3 Topic Modeling

The underlying basis of topic models is the straightforward idea that documents are collections of topics each of which represents a probability distribution over words. Finding the main topics in a corpus of texts that are meant to be thematically related, cohesive, and self-contained is the main objective of topic models [21]. The most popular topic models are probabilistic graphical models, in which topics are frequently represented as distributions over words and documents are typically represented as distributions over subjects. LDA [22] is a prime illustration. The underlying components are modeled as coordinate axes in NMF, a topic mining technique that varies from earlier probabilistic perspectives in that each document is equivalent to a distinct point in the latent linear space [23].

### 3.4 Sentiment Analysis

Sentiment analysis is a subset of natural language processing (NLP) that looks for user sentiments in reviews, comments, and other online content. Nowadays, user reviews on a wide range of subjects, including movies, news, food, fashion, politics, COVID-19, and much more, are often shared on social media platforms like Facebook, Twitter, and Reddit [24].

### 3.5 BERT

We present a sentiment analysis approach for WFH, also known as digital nomadism. To preprocess the data, we employ the Spacy and Gensim libraries, which offer powerful tools for tasks such as tokenization, lemmatization, and stopword removal. Next, we integrate BERT (Bidirectional Encoder Representations from Transformers), a state-of-the-art NLP model, into our analysis pipeline. BERT, which stands for Bidirectional Encoder Representations from Transformers, is pre-trained on a vast amount of unlabeled text and can be fine-tuned for various downstream tasks. We proceed to fine-tune the pre-trained BERT model on a labeled dataset specific to our sentiment analysis task, consisting of examples with their corresponding sentiment labels (positive, negative, or neutral). By training BERT on this dataset, it learns to accurately classify the sentiment of the working from home text. Leveraging BERT's capability to capture contextual information and semantic nuances, we achieve high accuracy in sentiment classification.

Finally, we evaluate the performance of BERT against other machine learning models, ultimately determining BERT to be the best-performing model for our sentiment analysis task.

## 4 Results and Discussion

### 4.1 Data Overview

Between December 2020 and August 2022, 14512 comments were collected from the 7 subreddits. Table 2 includes summary statistics for subreddits.

**Table 2.** Descriptive Statistics for Extracted comments.

Subreddit	Total number of comments
digitalnomad	6788
homeoffice	46
RemoteJobs	1036
comment	2024
workfromhome	1287
WorkOnline	2143

### 4.2 Sentiment Based Results

The goal of sentiment analysis, also known as opinion mining, is to locate and categorize the emotions (positive, negative, or neutral) associated with a text. It makes it simpler for decision-makers or authorities to recognize and evaluate people's opinions automatically, which is especially useful for serious issues. Given this information, one of the objectives of this study is to develop a high-performing sentiment analysis that can accurately determine the sentiment of Reddit comments on WFH-related topics. We investigated various conventional machine methods, deep learning models, and contextualized word embedding models for sentiment analysis after labeling the dataset. Results of deep learning models and conventional machine methods for sentiment analysis are shown in Table 3. Results of contextualized word embedding models for sentiment analysis are displayed in Table 4. BERT has the best performance metrics among all models. Figure 1 displays the total number of favorable, unfavorable, and neutral remarks. When all comments are taken into account, the two favorable ones come out on top. It demonstrates that attitudes toward remote employment are favorable.

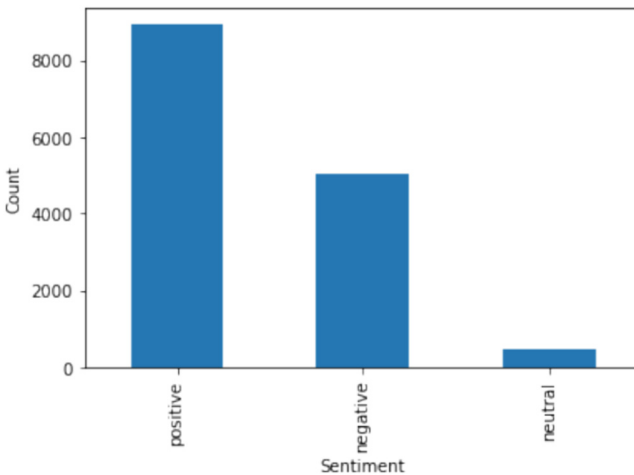


**Table 3.** Performance results of traditional machine learning techniques and deep learning models for sentiment analysis.

Methods	Accuracy	F-Measure	Precision	Recall
SVM	0.86	0.75	0.83	0.71
CNN (1D)	<b>0.88</b>	0.82	0.86	0.79
MLP	0.83	0.82	0.82	0.83
LSTM	0.84	0.82	0.81	0.84

**Table 4.** Performance results of contextualized word embedding models for sentiment analysis.

Methods	Accuracy	F-Measure	Precision	Recall
BERT	<b>0.93</b>	0.93	0.93	0.93
MBERT	0.62	0.47	0.38	0.62
RoBERTa	0.90	0.90	0.90	0.90
DistilBERT	0.93	0.93	0.93	0.93
GPT-2	0.85	0.85	0.85	0.85

**Fig. 1.** Total number of positive, neutral, and negative comments

### 4.3 Polarity Based Analysis

The NMF model generates 10 latent topics from each positive, negative, and neutral comment about WFH. The NMF approach has been seen to achieve worthwhile outcomes. First, when the topics obtained from positive comments are examined, it has been seen that visa, digital nomad, tax system, remote working policies, providing communication

security with remote working, software and hardware components required for remote working, and professions suitable for remote working are mentioned.

Today, it is the dream of many people, especially the younger generation, to continue life regardless of location. Moreover, it is now very easy to go from one end of the world to the other with the conveniences brought by globalization. As such, new business models attract the attention of the young audience. Digital nomadism, which is the trend concept of recent times, is one of these business models. In relation to this, it is expected that digital nomadism, digital nomad visa, visa, taxes, and professions suitable for remote working are the topics that are widely discussed in the user comments during the Covid-19 period. It has been seen those keywords under topic 1, topic 2, topic 6, topic 7, and topic 10 are related to these headings. When topic 10 is examined, it is seen that Croatia and Bali are frequently mentioned in the comments regarding digital nomadism. For example, a comment states “That forb site I assume fairly credible. This Indonesia does digital nomads pay tax Bali? The planned digital nomad visa valid years. It allows remote workers live country tax free long income derives businesses based outside Indonesia. The special permit designed simple solution foreigners wanting work Asian country Croatia Croatia said paradise digital nomads benefit total exemption taxes country”. The topics extracted from positive comments are given in Fig. 2.

Negative comments are grouped under the headings of remote work, digital nomad visa, digital nomadism, and digital nomad, and online job applications, security related to remote connection. It is seen that there are comments on the illegality of remote working. This is clearly seen in the comment “No reside country 6 months countries pay tax. Others lesser period. Not illegal. And yes, remote working tourist visa countries period 6 months staying illegal”. One of the comments about security related to remote connection in negative comments is “Does VPN bypass security concerns? As people thread saying remote work likely NOT allowed” is in the form. The topics extracted

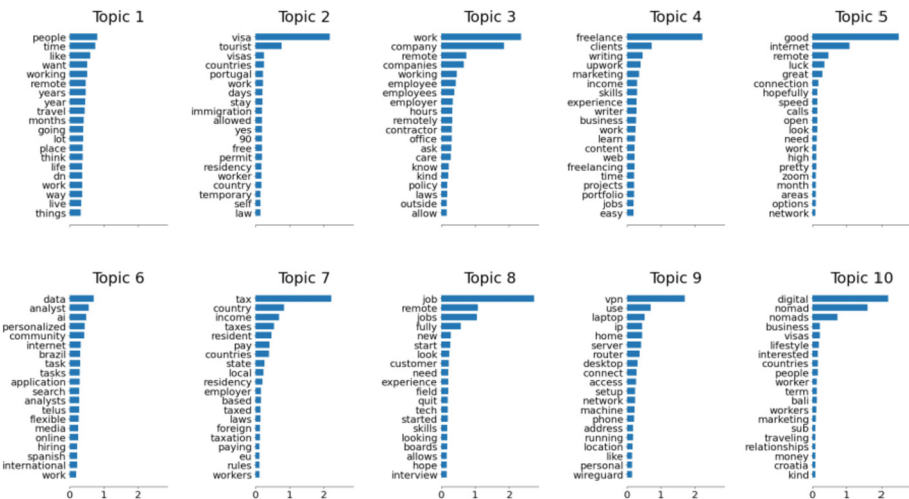


Fig. 2. Topics extracted from positive comments.

from positive comments are given in Fig. 3. When the topics extracted from the neutral comments are examined, it has seen that no pattern is found.

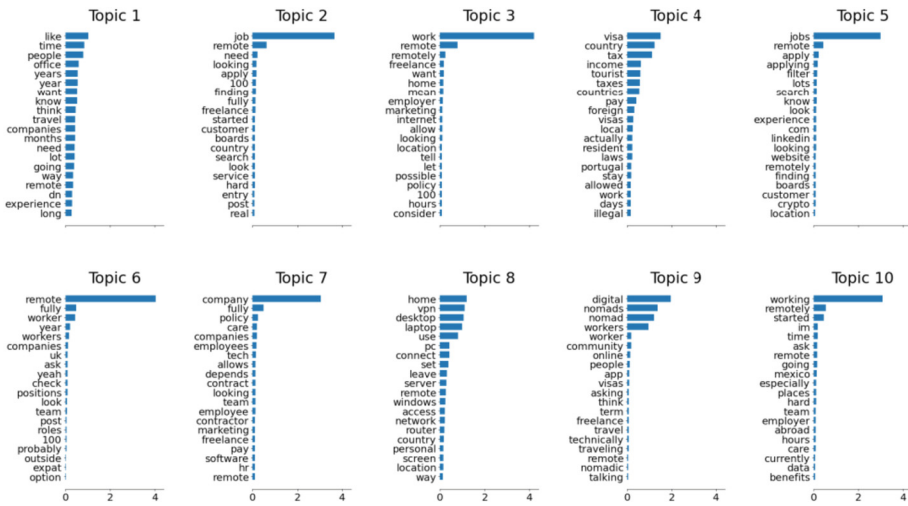


Fig. 3. Topics extracted from negative comments.

#### 4.4 Monthly Analysis

The monthly analysis covers the period from March 2021-July 2022. When all the months in the relevant period are examined separately, it is seen that digital nomadism, visa, and tax topics are definitely discussed every month. In this section, the topics discussed will be summarized on a monthly basis, except for the digital nomadism, visa, and tax topics.

When the user comments made in March and April 2021 are examined, it is seen that UNICEF was focused on. UNICEF has worked to alleviate the impact of Covid-19 on vulnerable children and families affected by this pandemic. A few of the keywords under this topic are fund, children, consultant, consultancy, international program and so on. Again, in April 2021, it is seen that there are comments that Mexico is a suitable route in the pandemic. For example, “thank i working moment jobs great partner permanently scored remote work ecstatic we stay places 2 4 months time usually sense place currently New Mexico amazing place pandemic looking forward abroad”. Digital nomadism, taxes, and jobs suitable for remote work are among the topics covered in April 2021. While taxes were still on the agenda in May 2021, software and hardware that will be required for remote working were also among the topics discussed. June 2021 appears to be a month when remote work is discussed. In July 2021, topics were obtained about companies hiring remote workers globally. One of the comments to illustrate this is “yeah tier companies hiring remote right, so software engineer product jobs remote wasn’t 18 months ago”. In August 2021, unlike in other months, it was observed that the topics related to remote copyediting were talked.

The healthcare industry has experienced a shift toward remote employment during the epidemic. Appointments made through phone or video chats allowed patients to receive the guidance they required in a secure and timely manner, and they are likely to continue to be an option in the future. This situation is one of the most talked about topics of October 2021. In November and December of 2021, comments were made on hybrid work. Comments made in the first month of 2022 are on digital marketing, especially visa and taxation for Portugal and Spain. In February 2022, comments were made on the remote recruitment of DevOps, software, python and java developers. To give an example from the comments, we can give "...top 5 freelance DevOps engineers scale team cost effectively timezone friendly...". In April 2022, it was observed that security issues were discussed in recruitment and remote working on LinkedIn. May 2020 has been a month in which professions suitable for remote working, designers and developers, remote workers insurance issues, remote working security and hybrid working issues were discussed. In June 2022 artificial intelligence and security issues were discussed. Finally, In July 2022, it was observed that the job application processes were talked.

#### **4.5 Discussion**

The majority of research on remote working has concentrated on comparing employees who engage in remote working and who do not, but little emphasis has been paid to the potential causes for these results. Understanding their thoughts and feelings is important because many employees were required to work remotely due to the COVID-19 pandemic. Reddit widespread use has given researchers the chance to investigate how the general public views remote working during the COVID-19 pandemic and to identify potential advantages and difficulties. Using sentiment analysis and NMF topic model, this study focuses on the public's attitude about working remotely, topic extraction under three different sentiment polarity (positive, negative, and neutral), and topics on a monthly basis. The headings related to digital nomadism, visa, and tax make up the majority of the extracted topics. Different topics show emotional differences, suggesting that user attitudes varied.

Experiments showed that the majority of tweets relating to remote working had somewhat positive opinions. In the positive comments, it was observed that the good aspects of remote working and the digital nomad were emphasized, while the negative comments mentioned illegality and security problems. It was observed that the low number of neutral comments prevented a pattern from being extracted from these comments. When we look at the monthly analyzes, it is clearly seen that certain topics are definitely discussed every month. Apart from this, it has been understood that the topics that change on a monthly basis are determined according to the agenda of that month.

## **5 Conclusion and Future Works**

By type of work, several remote working habits exist. Early on, in the 1980s, when home-based digital technology was still in its infancy, the characteristics of certain jobs—jobs that required little more than a telephone and a computer to perform—had high levels

of individual control over the work, clear deadlines and deliverables, required concentration, and little need for communication. These qualities still hold true for freelance employment and remote work today. Strong evidence of how remote employment varies by occupation has emerged more recently with the impact of the COVID-19 pandemic. Managers, financial analysts, engineers, computer programmers, and lawyers were discovered to be among those whose professions were performed in a range of settings, including at home. It's also important to remember that remote working trends preceded the pandemic. People were obliged to WFH in an unanticipated manner as a result of the COVID-19 pandemic, which caused a noticeable jolt to our usual working patterns. The majority of hiring managers claim that as a result of COVID-19, their staff will be more distributed in the future.

Together, topic mining and sentiment polarity analysis may accurately reflect the themes and opinions Reddit people have on WFH. This study offers fresh viewpoints for sensing the WFH inside the COVID-19 and behavioral analytics framework. Overall, our results demonstrate that social media data may be used to more effectively explain pandemic challenges and attitudes toward WFH trends, enabling more individualized and publically just policies. This study assists managers in comprehending how WFH is seen both before and after the pandemic.

Although the analysis formed the basis for the results, there are a number of limitations that might motivate further study. Our findings were supported by the reliability of the information discovered using our search terms. We looked for postings mentioning the WFH on Reddit, and we discovered text snippets that were typical of WFH perceptions. We are certain that our data is accurate as a consequence. Any use of Reddit data comes with a number of limitations and drawbacks. Reddit users' demographic makeup is unknown because no individual data is collected. Some data suggests that compared to the general population, Reddit users are more likely to be male and younger. Our results should be interpreted in light of the likely age and gender skew in our data. We could have missed any arguments or concerns that were raised but unreported in our study. So, in order to get a more accurate picture of the conceptual aspects of the WFH discussion on Reddit, we should expand the scope of data with other social media platforms in future works. Reddit user interactions, such as upvotes and downvotes, may help us better understand current subjects, but we choose not to include them in this study. Future works can accomplish this.

## References

1. Ceja, E.G.: Behavior Analysis with Machine Learning Using R. CRC Press, London (2022)
2. Kawsar, F., Min, C., Mathur, A., Montanari, A.: Earables for personal-scale behavior analytics. *IEEE Pervasive Comput.* **17**, 83–89 (2018)
3. Pinola, M.: The 7 biggest remote work challenges (and how to overcome them). <https://zapier.com/blog/remote-work-challenges/>
4. Eken, S.: A topic-based hierarchical publish/subscribe messaging middleware for COVID-19 detection in X-ray image and its metadata. *Soft. Comput.* **27**, 11073 (2020). <https://doi.org/10.1007/s00500-020-05387-5>
5. Allen, J., Reed, K.: Suddenly Virtual Making Remote Meetings Work. Wiley, Hoboken (2021)
6. Althoff, L., Eckert, F., Ganapati, S., Walsh, C.: The geography of remote work (2021)

7. Alipour, J.-V., Fadinger, H., Schymik, J.: My home is my castle – the benefits of working from home during a pandemic crisis. *J. Public Econ.* **196**, 104373 (2021)
8. Baruch, Y.: Teleworking: Benefits and pitfalls as perceived by professionals and managers. *N. Technol. Work. Employ.* **15**, 34–49 (2000)
9. Ahmad, T.: Corona virus (COVID-19) pandemic and work from home: challenges of cybercrimes and cybersecurity. *SSRN Electron. J.* (2020)
10. Madero Gómez, S., Ortiz Mendoza, O.E., Ramírez, J., Olivás-Luján, M.R.: Stress and myths related to the COVID-19 pandemic's effects on remote work. *Manag. Res. J. Iberoamerican Acad. Manag.* **18**, 401–420 (2020)
11. Wang, B., Liu, Y., Qian, J., Parker, S.K.: Achieving effective remote working during the COVID-19 pandemic: a work design perspective. *Appl. Psychol.* **70**, 16–59 (2020)
12. Cifuentes-Faura, J.: The importance of behavioral economics during COVID-19. *J. Econ. Behav. Stud.* **12**, 70–74 (2020)
13. Caligiuri, P., De Cieri, H., Minbaeva, D., Verbeke, A., Zimmermann, A.: International HRM Insights for navigating the COVID-19 pandemic: Implications for future research and practice. *J. Int. Bus. Stud.* **51**, 697–713 (2020)
14. Leonardi, P.M.: Covid-19 and the new technologies of organizing: digital exhaust, digital footprints, and artificial intelligence in the wake of remote work. *J. Manag. Stud.* **58**, 249–253 (2020)
15. Wrycza, S., Maślankowski, J.: Social media users' opinions on remote work during the COVID-19 pandemic. Thematic and sentiment analysis. *Inf. Syst. Manag.* **37**, 288–297 (2020)
16. Dubey, A.D., Tripathi, S.: Analysing the sentiments towards work-from-home experience during COVID-19 pandemic. *J. Innov. Manag.* **8**, 13–19 (2020)
17. Zhang, C., Yu, M.C., Marin, S.: Exploring public sentiment on enforced remote work during COVID-19. *J. Appl. Psychol.* **106**, 797–810 (2021)
18. Daneshfar, Z., Asokan-Ajitha, A., Sharma, P., Malik, A.: Work-from-home (WFH) during COVID-19 pandemic – a NETNOGRAPHIC investigation using Twitter data. *Inf. Technol. People* (2022)
19. Emoji. <https://pypi.org/project/emoji/>
20. Gensim: Topic modelling for humans. <https://radimrehurek.com/gensim/>
21. Ekinci, E., İlhan Omurca, S.: NET-LDA: a novel topic modeling method based on semantic document similarity. *Turkish J. Electr. Eng. Comput. Sci.* **28**, 2244–2260 (2020)
22. Ekinci, E., İlhan Omurca, S.: Concept-LDA: incorporating babelify into LDA for aspect extraction. *J. Inf. Sci.* **46**, 406–418 (2019)
23. Chen, Y., Zhang, H., Liu, R., Ye, Z., Lin, J.: Experimental explorations on short text topic mining between LDA and NMF based schemes. *Knowl.-Based Syst.* **163**, 1–13 (2019)
24. Yurtsever, M.M., Shiraz, M., Ekinci, E., Eken, S.: Comparing covid-19 vaccine passports attitudes across countries by analysing Reddit comments. *J. Inf. Sci.* 016555152211483 (2023)



# An Improved Data Classification in Edge Cloud-Assisted IoMT: Leveraging Machine Learning and Feature Selection

Abdelkarim Ait Temghart<sup>1(✉)</sup>, Mbarek Marwan<sup>2</sup>, and Mohamed Baslam<sup>1</sup>

<sup>1</sup> TIAD Laboratory, FST, Sultan Moulay Slimane University, Beni Mellal, Morocco  
aittemghart.abdelkarim@gmail.com

<sup>2</sup> ENSIAS, Mohammed V University, Rabat, Morocco

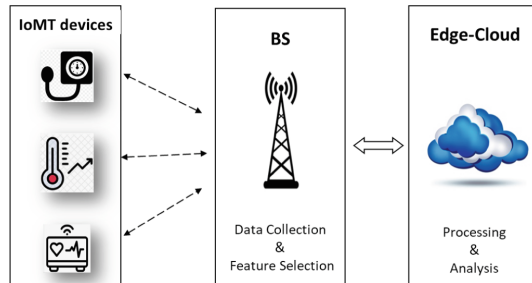
**Abstract.** As data generated by the IoMT devices are offloaded to the cloud, processing and analyzing such large-scale data to extract useful information presents challenges in terms of latency, bandwidth, and privacy. Edge computing has emerged as a promising paradigm to address these challenges. For this purpose, the proposed approach explores the benefits of performing Feature Selection methods, including filters, wrappers, and embedded techniques, along with ML algorithms at the edge. As a result, a novel framework is proposed that requires an optimal number of features from the dataset to build an optimal data classification model. The results of the study show that feature selection can significantly improve the classification accuracy and performance of ML algorithms when applied to the medical dataset. Moreover, the simulation results confirm that the XGBoost classifier utilizing the ET algorithm achieved the highest classification accuracy of 95% surpassing the current state-of-the-art.

**Keywords:** Classification · Feature Selection · Machine learning · XGBoost Classifier · Extra Tree · Network Attacks · IoMT

## 1 Introduction

It is evident that there is a pressing need for innovative security mechanisms that can effectively and efficiently address the security challenges facing the Internet of Medical Things (IoMT) networks [1, 5]. Traditional IDS architectures have primarily relied on centralized deployment in data centers, often resulting in high latency and increased network traffic [18]. Taking a step in this direction, Edge-Cloud computing [13] is seen as a promising security solution in both industry and research, believed to efficiently protect IoMT networks. Therefore, the extensive variety of possible attacks and their evolution makes the issue more difficult to solve. Edge cloud leverages the computational resources and storage capabilities of edge devices, such as routers, gateways, and access points, to process and analyze data closer to the source. To address this concern, as shown in

Fig. 1, we proposed a system for feature selection in edge cloud. Its purpose is to ensure that only the most relevant information is processed and analyzed, reducing the computational burden and mitigating the risk of data breaches and also performing using various techniques of Machine Learning. We start by describing the situation and doing a state-of-the-art study of this security system for IoMT networks. We take into account widely available open datasets and use the WUSTL-EHMS-2020 dataset in order to test and evaluate solutions. We put many ML algorithms [33] to the test and illustrate that it is actually simple to get consistent results by following the approach mentioned in Fig. 2. One of the key factors that affect the performance of these algorithms is the selection of features from the dataset. Which is the process of identifying and selecting a subset of relevant features from a larger set of features with the goal of improving the classification accuracy and performance of ML algorithms. There are various feature selection techniques that have been proposed, including wrapper methods, filter methods, and embedded methods. Wrapper methods use the performance of a specific machine learning algorithm as the evaluation criterion for feature selection. Filter methods, on the other hand, use statistical measures to evaluate the relevance of features. Embedded methods combine feature selection and learning into a single process.

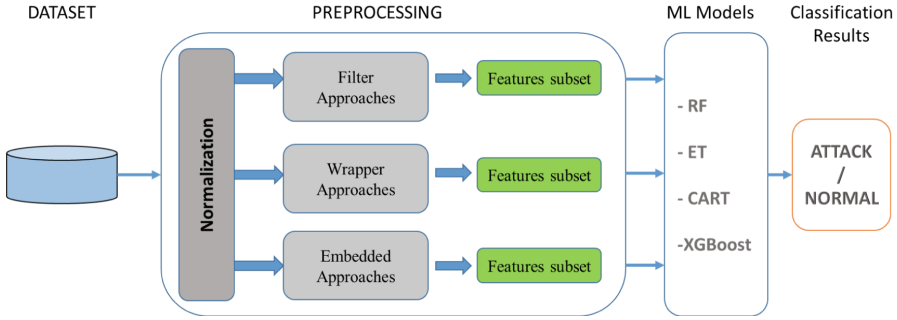


**Fig. 1.** Flowchart of the proposed methodology.

In this study, we make several contributions to the fields of machine learning and medical applications. Some of these contributions include:

- 1) Investigating the effect of feature selection on the classification accuracy and performance of machine learning algorithms using a medical dataset.
- 2) Applying different feature selection techniques to a medical dataset and comparing their performance.
- 3) Testing the performance of multiple machine learning algorithms (Decision Trees, Random Forest, XGBoost, and Extra randomized Tree) after feature selection.
- 4) A novel framework that selects an optimal number of features for building an optimal data classification model.





**Fig. 2.** Proposed Framework.

The rest of the paper is organized as follows: Sect. 2 presents recent work concerning existing security threats in the IoMT. Section 3 provides a discussion on the machine learning methods and techniques that have been used to ensure their efficiency on the suggested datasets, and Sect. 4 presents the evaluation and experimental results of our proposed approach. Section 5 concludes with the conclusion and future directions.

## 2 Literature Review

Machine learning has been applied in IoMT to analyze the data generated by these devices [22], but there are challenges in dealing with the large amount of data generated by medical IoMT devices. Several studies have been done to mitigate the vulnerabilities of attack on IoMT in reason to improve the performance of this smart system and to reduce the involvement of human [9, 12, 16, 21]. In machine learning-based models, the feature selection approach is regarded as the most crucial stage. A great deal of research has been done on Feature Selection Techniques (FS) that use classification methods. Previous studies have investigated the impact of feature selection on the classification accuracy and performance of machine learning algorithms when applied to medical datasets [7, 17, 27]. the author of [23] describes a feature selection algorithm using mutual information for cancer microarray data. The authors proposed a feature selection method based on qualitative mutual information, which incorporates the relationship between genes and samples in the data. They applied their method to several cancer microarray datasets and demonstrated that their method is effective in identifying important genes for cancer classification. Zhang et al. in [32] introduced WRHFS, a feature selection technique that determines the heart stroke risk efficiently. WRHFS uses filter-based methods such as Fisher score, information gain, and standard deviation to weigh and rank features. It selects crucial input features for heart stroke prediction based on available knowledge. Singh et al. [29] proposed a stroke prediction method using the CHS dataset. They employed DT for feature selection, followed by reducing the feature space

dimensionality through PCA. The classification model was built using an MLP network. In [10], a Genetic Algorithm (GA) wrapper was used to detect heart disease by selecting the most significant features. The algorithm selected key features from the Cleveland heart disease dataset for heart disease detection. The final features were then used to evaluate the accuracy with a Support Vector Machine (SVM). In [26], The optimal feature subset was obtained using the chi-squared, ReliefF, and correlation-based feature subset evaluator. The highest accuracy of 86.46% was achieved by the SMO classifier when using the optimal feature set selected by the chi-squared feature selection method. Alzagebah et al. in [4] proposed a modified wrapper feature selection method that uses MFO algorithm and neighborhood search to select the most informative features. The modified MFO algorithm has shown superior performance when compared to other algorithms such as particle firefly algorithm (FFA), genetic algorithms (GA), and swarm optimization (PSO). In [28], The authors employed Particle Swarm Optimization (PSO) to select features in the dataset. and machine learning/deep learning (ML/DL)-based models to detect cyber attacks in IoMT. The authors evaluated the system using the NSL-KDD datasets and found that the PSO and random forest (RF)-based solution achieved an accuracy of 99.76%. In [15] the author uses the Minimum-maximum concept of normalization to feature scaling on the UNSW-NB15 dataset. Moreover, the dimensionality reduction was performed by PCA (Principal Component Analysis). However, there are limited studies that investigate the impact of feature selection on machine learning algorithms when applied to medical IoMT datasets. In conclusion, the literature review highlights the importance of feature selection in the context of medical applications and the potential impact of irrelevant features on the performance of ML algorithms. It also highlights the potential of Medical IoMT to improve patient care and the need for more research in this area.

### 3 Methodology

As irrelevant features can decrease accuracy, this research relies on feature selection techniques in improving learning performance and lower computational cost in Edge cloud, especially for the datasets with many variables and features. In the same line, it is crucial to choose the right machine-learning model. Practically, we perform classification using different ML models with and without feature selection. Additionally, there are various statistical estimation methods and metrics that determine the most appropriate method. Ultimately, we determine the optimal number of features required to construct an optimal data classification model.

#### 3.1 Dataset

The paper evaluates a data model using the ‘wustl-ehms’ IoMT dataset [12], which was derived from a real-time health monitoring system. A dataset comprising both patient biometric and network traffic data was generated using the

ARGUS tool. The biometric data includes measures such as pulse rate, temperature, blood pressure, oxygen saturation, heart rate, respiration rate, etc. In addition, network traffic flow records and their metrics were also collected to obtain features describing the overall network traffic. In total, the dataset includes 44 features, 34 of which are related to network traffic. The dataset is labeled as either an attack or normal traffic, with attacks labeled as “0” and normal traffic labeled as “1”. Table 1 provides a description of each of the features in the dataset, as well as the feature type.

**Table 1.** Dataset Features.

Metric	Description
SIntPktAct	Source active inter packet
Packet_num	number of packet
DIntPktAct	Destination active inter packet
SrcJitter	Source jitter
DstJitter	Destination jitter
sMaxPktSz	Source maximum transmitted packet size
dMaxPktSz	Destination maximum transmitted packet size
sMinPktSz	Source minimum transmitted packet size
dMinPktSz	Destination minimum transmitted packet size
Dur	Duration
Trans	Aggregated packets count
TotPkts	Total packets count
TotBytes	Total packets bytes
Loss	Retransmitted or dropped packets
pLoss	Percentage of retransmitted or dropped packets
pSrcLoss	Percentage of source retransmitted or dropped packets
pDstLoss	Percentage of destination retransmitted or dropped packets
Rate	Number of packets per second
Load	Load
Temp	Temperature
SpO2	Peripheral oxygen saturation
Pulse_Rate	Pulse rate
SYS	Systolic blood pressure
DIA	Diastolic blood pressure
Heart_Rate	Heart rate
Resp_Rate	Respiration rate
ST	ECG ST segment

### 3.2 Feature Selection

We applied three different feature selection techniques to the dataset including filter, embedded, and wrapper methods [19]. Filter Method uses a statistical measure to evaluate the importance of each feature independently of the model. An example of a filter method is the Chi-squared test [25], which measures the dependence between two variables and can be used to select features that are most informative with respect to the target variable. We propose to use in addition to chi-squared, Mutual Information (MI). [2, 6] and Principle Component Analysis (PCA) [30] as filter methods. In addition to the filter method, another widely used feature selection is the Embedded Method. This method combines feature selection and model training into one process. An example of

an embedded method is XGBoost [3]. Another popular feature selection technique is the wrapper method, which evaluates the performance of a subset of features by training a machine learning model on them. This method is computationally expensive because it requires training a model for each subset of features. PSO [24] and GWO [8] are two examples proposed for this method. It's important to try different methods and evaluate their performance. Additionally, it's important to consider the trade-off between the number of features and the performance of the model.

### 3.3 Classification Methods

In order to ensure that our methodologies employed have to prove a great performance in the detection of IDSs and have to be evaluated successfully in various models. Among models used in This study are Classification and Regression Trees (CART), Random Forest (RF), Extra Tree (ET) [31], and XGBoost techniques [14]. We used the scikit-learn library to implement these algorithms. We used default parameters for all the models and did not perform any hyperparameter tuning. To evaluate the performance of machine learning models on a given dataset, we use a variety of metrics, including accuracy, precision, recall, and F-score [20]. These metrics are important for evaluating the chosen machine learning models, as they provide different insights into the model's performance. These metrics are used to provide a comprehensive understanding of the model's performance on the dataset.

## 4 Results and Discussion

This section presents the results of the study, including the results of applying different FS techniques on the dataset and the effect on the classification accuracy and performance of the ML algorithms. Table 2 presents the results of various feature selection (FS) techniques applied to the medical dataset. The Particle Swarm Optimization (PSO) and Grey Wolf Optimization (GWO) feature selection methods both achieved an accuracy of 94% when utilized with a decision tree algorithm by selecting subsets of 11 and 8 features respectively. Meanwhile, the Principal Component Analysis (PCA) feature selection method selected 18 features with a 93.30% accuracy when used with a Random Forest algorithm. The Chi-Squared method showed a 94.81% accuracy with the Extremely Randomized Trees (ET) algorithm. The XGBoost method achieved a 95.28% accuracy by selecting a subset of 13 features with the ET algorithm. Finally, the Mutual Information method demonstrated 94.60% accuracy by selecting 10 features and utilizing the Classification and Regression Tree (CART) algorithm.

The results of our study indicate that feature selection has a significant impact on the classification accuracy and performance of machine learning algorithms. XGBoost feature selection method performed the best with the ET algorithm, while the PCA method performed the best with the RF algorithm. And the mutual information feature selection method performed the best with

**Table 2.** Features selected size and Accuracy.

Feature Selection Technique	Feature size	RF	ET	CART	XGBoost
All features	34	<b>92.81</b>	92.29	92.68	92.80
PSO	11	94.19	<b>94.73</b>	93.21	93.36
GWO	8	93.38	<b>94.19</b>	92.89	92.95
PCA	18	<b>93.30</b>	93.16	92.28	93.01
CHI-Squared	22	93.30	<b>94.81</b>	92.64	93.21
MI	10	92.93	92.13	<b>94.60</b>	93.33
XGBoost	13	94.17	<b>95.28</b>	92.50	93.34

**Table 3.** Machine learning methods using XGBoost feature selection.

Feature Selection	ML models	Accuracy	Precision	Recall	Fscore
XGBOOST	<i>RF</i>	94,17	94,28	55,86	70,64
	<i>ET</i>	<b>95,28</b>	95,26	65,96	77,80
	<i>CART</i>	92,50	92,54	70,68	70,28
	<i>XGB</i>	93,34	93,70	47,55	64,17

**Table 4.** A comparison of our work with the prior art.

Article	ML models	Accuracy	limit
[12]	KNN	90%	performance can be improved
[11]	Tree classifier	93%	Data augmentation results in unrealistic dataset
Our work	XGBoost classifier	95%	performance can be improved using Deep learning

the decision tree algorithm (CART). This is consistent with previous research on feature selection, which has shown that different feature selection methods may be more suitable for different machine learning algorithms.

Our study also shows that feature selection improves the performance of machine learning algorithms compared to using all of the features. This is an important finding, as it highlights the importance of feature selection in machine learning. In addition, Table 3 shows that the Extreme Gradient Boosting method and the Extremely Randomized Trees (ET) algorithm resulted in a high accuracy of 95%. Moreover, Fig. 3 demonstrates that the optimal number of features selected to achieve this accuracy was 13. This is an interesting finding, as ET is known for its fast training times and good performance on high-dimensional data and datasets with a large number of features.

Furthermore, We achieved 95% accuracy, surpassing the current state-of-the-art as shown in Table 4. Finally, the results suggest that there is a trade-off between the number of selected features and the performance of the model and it's important to consider this trade-off when selecting the features.

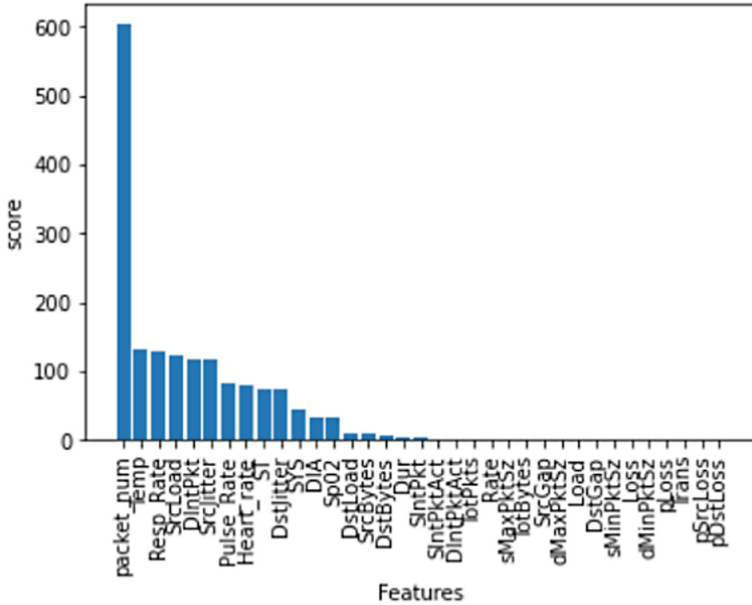


Fig. 3. XGBoost feature selection.

## 5 Conclusion

In summary, our study has investigated the effect of feature selection on the classification accuracy and performance of machine learning algorithms using a medical dataset. We have applied various feature selection techniques such as wrapper (PSO, GWO). Filter (MI, PCA, CHI-squared), and embedded (XGBoost). Our results indicate that feature selection has a significant impact on the classification accuracy and performance of machine learning algorithms using a medical dataset (WUSTL-EHMS). The XGBoost feature selection method outperformed others when combined with the ET algorithm. We also observed a trade-off between the number of selected features and model's performance. XGBoost achieved the highest classification accuracy of 95% with an optimal selection of 13 features, making it the preferred choice for our classification model. Additionally, For future prospects, we propose combining feature selection techniques with deep learning algorithms to enhance the performance of classification.

## References

1. Ait Temghart, A., Outanoute, M., Marwan, M.: Game theoretic approaches to mitigate cloud security risks: an initial insight. In: Fakir, M., Baslam, M., El Ayachi, R. (eds.) CBI 2021. LNBP, vol. 416, pp. 335–347. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-76508-8\\_24](https://doi.org/10.1007/978-3-030-76508-8_24)

2. Alhaj, T.A., Siraj, M.M., Zainal, A., Elshoush, H.T., Elhaj, F.: Feature selection using information gain for improved structural-based alert correlation. *PLoS ONE* **11**(11), e0166017 (2016)
3. Alsahaf, A., Petkov, N., Shenoy, V., Azzopardi, G.: A framework for feature selection through boosting. *Expert Syst. Appl.* **187**, 115895 (2022)
4. Alzaqebah, M., Alrefai, N., Ahmed, E.A.E., Jawarneh, S., Alsmadi, M.K.: Neighborhood search methods with moth optimization algorithm as a wrapper method for feature selection problems. *Int. J. Electr. Comput. Eng. (IJECE)* **10**(4), 3672 (2020)
5. Aziz, M.F., Khan, A.N., Shuja, J., Khan, I.A., Khan, F.G., Khan, A.U.R.: A lightweight and compromise-resilient authentication scheme for IoTs. *Trans. Emerg. Telecommun. Technol.* **33**(3) (2022)
6. Beraha, M., Metelli, A.M., Papini, M., Tirinzoni, A., Restelli, M.: Feature selection via mutual information: new theoretical insights (2019). Publisher: arXiv Version Number: 1
7. Chaganti, R., Mourade, A., Ravi, V., Vemprala, N., Dua, A., Bhushan, B.: A particle swarm optimization and deep learning approach for intrusion detection system in internet of medical things. *Sustainability* **14**(19), 12828 (2022)
8. Emary, E., Zawbaa, H.M., Grosan, C., Hassenian, A.E.: Feature subset selection approach by gray-wolf optimization. In: Abraham, A., Krömer, P., Snaes, V. (eds.) *Afro-European Conference for Industrial Advancement. AISC*, vol. 334, pp. 1–13. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-13572-4\\_1](https://doi.org/10.1007/978-3-319-13572-4_1)
9. Falco, G., Caldera, C., Shrobe, H.: IIoT cybersecurity risk modeling for SCADA systems. *IEEE Internet Things J.* **5**(6), 4486–4495 (2018)
10. Gokulnath, C.B., Shantharajah, S.P.: An optimized feature selection based on genetic approach and support vector machine for heart disease. *Clust. Comput.* **22**(S6), 14777–14787 (2019)
11. Gupta, K., Sharma, D.K., Datta Gupta, K., Kumar, A.: A tree classifier based network intrusion detection model for internet of medical things. *Comput. Electr. Eng.* **102**, 108158 (2022)
12. Hady, A.A., Ghubaish, A., Salman, T., Unal, D., Jain, R.: Intrusion detection system for healthcare systems using medical and network data: a comparison study. *IEEE Access* **8**, 106576–106584 (2020)
13. Haseeb-Ur-Rehman, R.M.A., et al.: Sensor cloud frameworks: state-of-the-art, taxonomy, and research issues. *IEEE Sens. J.* **21**(20), 22347–22370 (2021)
14. Jiang, H., He, Z., Ye, G., Zhang, H.: Network intrusion detection based on PSO-XGBoost model. *IEEE Access* **8**, 58392–58401 (2020)
15. Kayode Saheed, Y., Idris Abiodun, A., Misra, S., Kristiansen Holone, M., Colomo-Palacios, R.: A machine learning-based intrusion detection for detecting internet of things network attacks. *Alex. Eng. J.* **61**(12), 9395–9409 (2022)
16. Kim, S., Park, K.J.: A survey on machine-learning based security design for cyber-physical systems. *Appl. Sci.* **11**(12), 5458 (2021)
17. Kumar, P., Gupta, G.P., Tripathi, R.: An ensemble learning and fog-cloud architecture-driven cyber-attack detection framework for IoMT networks. *Comput. Commun.* **166**, 110–124 (2021)
18. Marwan, M., Temghart, A.A., Sifou, F., AlShahwan, F.: A cloud solution for securing medical image storage. *J. Inf. Organ. Sci.* **44**(1), 113–139 (2020)
19. Miao, J., Niu, L.: A survey on feature selection. *Procedia Comput. Sci.* **91**, 919–926 (2016)
20. Moustafa, N., Hu, J., Slay, J.: A holistic review of network anomaly detection systems: a comprehensive survey. *J. Netw. Comput. Appl.* **128**, 33–55 (2019)

21. Mujeeb Ahmed, C., Umer, M.A., Binte Liyakkathali, B.S.S., Jilani, M.T., Zhou, J.: Machine learning for CPS security: applications, challenges and recommendations. In: Maleh, Y., Shojafar, M., Alazab, M., Baddi, Y. (eds.) *Machine Intelligence and Big Data Analytics for Cybersecurity Applications*. SCI, vol. 919, pp. 397–421. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-57024-8\\_18](https://doi.org/10.1007/978-3-030-57024-8_18)
22. Mutlag, A.A., Abd Ghani, M.K., Arunkumar, N., Mohammed, M.A., Mohd, O.: Enabling technologies for fog computing in healthcare IoT systems. *Futur. Gener. Comput. Syst.* **90**, 62–78 (2019)
23. Nagpal, A., Singh, V.: A feature selection algorithm based on qualitative mutual information for cancer microarray data. *Procedia Comput. Sci.* **132**, 244–252 (2018)
24. Poli, R., Kennedy, J., Blackwell, T.: Particle swarm optimization: an overview. *Swarm Intell.* **1**(1), 33–57 (2007)
25. Ray, S., Alshouiliy, K., Roy, A., AlGhamdi, A., Agrawal, D.P.: Chi-squared based feature selection for stroke prediction using AzureML. In: *2020 Intermountain Engineering. Technology and Computing (IETC)*, pp. 1–6. IEEE, Orem (2020)
26. Reddy, K.V.V., Elamvazuthi, I., Aziz, A.A., Paramasivam, S., Chua, H.N., Prananand, S.: Heart disease risk prediction using machine learning classifiers with attribute evaluators. *Appl. Sci.* **11**(18), 8352 (2021). <https://doi.org/10.3390/app11188352>
27. Saba, T.: Intrusion detection in smart city hospitals using ensemble classifiers. In: *2020 13th International Conference on Developments in eSystems Engineering (DeSE)*, pp. 418–422. IEEE, Liverpool (2020)
28. Saheed, Y.K., Arowolo, M.O.: Efficient cyber attack detection on the internet of medical things-smart environment based on deep recurrent neural network and machine learning algorithms. *IEEE Access* **9**, 161546–161554 (2021)
29. Singh, M.S., Choudhary, P.: Stroke prediction using artificial intelligence. In: *2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*, pp. 158–161. IEEE, Bangkok (2017)
30. Song, F., Guo, Z., Mei, D.: Feature selection using principal component analysis. In: *2010 International Conference on System Science. Engineering Design and Manufacturing Informatization*, pp. 27–30. IEEE, Yichang (2010)
31. Wu, X., et al.: Top 10 algorithms in data mining. *Knowl. Inf. Syst.* **14**(1), 1–37 (2008)
32. Zhang, Y., Zhou, Y., Zhang, D., Song, W.: A stroke risk detection: improving hybrid feature selection method. *J. Med. Internet Res.* **21**(4), e12437 (2019)
33. Zolanvari, M., Teixeira, M.A., Gupta, L., Khan, K.M., Jain, R.: Machine learning-based network vulnerability analysis of industrial internet of things. *IEEE Internet Things J.* **6**(4), 6822–6834 (2019)





# Predicting Delays in Indian Lower Courts Using AutoML and Decision Forests

Mohit Bhatnagar<sup>1</sup>(✉) and Shivaraj Huchhanavar<sup>2</sup>

<sup>1</sup> Jindal Global Business School, OP Jindal Global University, Sonipat, India  
mohit.bhatnagar@jgu.edu.in

<sup>2</sup> Jindal Global Law School, OP Jindal Global University, Sonipat, India  
shivaraj.huchhanavar@jgu.edu.in

**Abstract.** This paper presents a classification model that predicts delays in Indian lower courts based on case information available at filing. The model is built on a dataset of 4.2 million court cases filed in 2010 and their outcomes over a 10-year period. The data set is drawn from 7000+ lower courts in India. The authors employed AutoML to develop a multi-class classification model over all periods of pendency and then used binary decision forest classifiers to improve predictive accuracy for the classification of delays. The best model achieved an accuracy of 81.4%, and the precision, recall, and F1 were found to be 0.81. The study demonstrates the feasibility of AI models for predicting delays in Indian courts, based on relevant data points such as jurisdiction, court, judge, subject, and the parties involved. The paper also discusses the results in light of relevant literature and suggests areas for improvement and future research. The authors have made the dataset and Python code files used for the analysis available for further research in the crucial and contemporary field of Indian judicial reform.

**Keywords:** Legal analytics · Judgement Delay Prediction · Pendency · AutoML · Decision Trees · XGBoost · Random Forests

## 1 Introduction

The expense, delay, and pendency have been enduring challenges facing the Indian courts [1, 2]. Studies have underlined various systemic inadequacies that affect the celerity, cost-effectiveness, and administrative efficiency of courts in India [1, 2]. Technology is seen as a panacea in optimizing judicial processes and case management. The judiciary, since 2005, has worked to build a robust computing infrastructure in courts across India [3]. To facilitate data-driven case and court management, the judiciary has also developed the National Judicial Data Grid [4]. However, the problem of pendency has not eased; on the contrary, pendency has increased significantly in recent years [5].

Against this backdrop, this paper demonstrates the potential for an AI model to predict the delay of court cases, using information such as the type of cases, and names of judges, advocates and parties. This predictive analysis would not only help litigants, advocates, and judges make informed decisions, but it would also help identify systemic

issues affecting the efficiency of courts and judges. This research is, to some extent, a departure from research that aims to predict or qualitatively assess judicial outcomes. Although the latter applications of machine learning and AI are no less relevant; however, India's judiciary faces elementary challenges that can only be addressed by improving the efficiency of judicial actors and institutions, and predictive modelling of the duration and delays of cases would be a step forward in this direction.

## 2 Setting the Context

AI and machine learning could strengthen judicial administration in various ways; the range of applications can encompass intelligent analytics, research and computational tools, with the eventual goal of implementing predictive justice [6]. These tools have the potential to provide comprehensive legal briefs on cases, including legal research and identifying crucial points of law and facts; thus, expediting the judicial process. Predictive AI tools can complement human judgment in the adjudicative process [7]. AI tools can also enhance the administrative efficiency of the courts and judges. It can facilitate e-filing, filtering, listing, notifying and prioritization of cases [8]. Additionally, intelligent tools, such as legal bots, can be designed to assist (potential) litigants in making informed decisions regarding their legal rights and to provide cost-effective access to basic legal services. Likewise, as this paper demonstrates, AI can also help predict the delays for a case in court.

For the purpose of this study, we see extant literature into two categories. Firstly, literature that endeavours to predict court case outcomes through artificial intelligence, and secondly, literature highlighting the effectiveness of AI in predicting court case delays. Over the past two decades, scholars have made significant attempts, resulting in noteworthy successes, to predict the outcomes of different types of court cases in several jurisdictions, notably China [9], the United States [10], the European Union [11], France [12], the United Kingdom, Brazil [13, 14] and India [15].

It is pertinent to note that predictive AI models have been employed mostly for outcome identification, outcome-based judgement categorisation, and outcome forecasting [16]. Identification of verdict in a full-text of a judgement is called outcome identification; whereas judgement categorisation is the task of categorising documents, based on the outcome. Likewise, as the term suggests, outcome forecasting involves the prediction of future decisions of courts [17]. However, so far, there has been no emphasis on predicting pendency in courts, this is true for India and elsewhere.


The Indian judiciary is keen on adopting AI tools in judicial administration. The judiciary has demonstrated foresight by being an early adopter of technology and making progress, albeit limited, in exploiting the capabilities of AI and data-driven insights. Advancements in ICT-enabled judicial administration are attained through the eCourts Mission Mode Project; the project set up a foundation for e-courts, equipping courts with fundamental computing hardware. In recent years, the Indian judiciary has taken a notable step to harness the potential of AI. In 2019, the then Chief Justice of India, Justice Bobde, launched the beta version of a neural translation tool called SUVAS [18]. This software can translate judicial rulings in English to nine vernacular scripts, and vice-versa. In 2021, the Supreme Court of India launched an AI portal: SUPACE, which

aims to utilize machine learning to manage large amounts of case data [19]. The Court has a dedicated Artificial Intelligence Committee, headed by a Supreme Court judge, to explore the use of AI in judicial administration [20].

In addition, with the support of the government, the judiciary has built the National Judicial Data Grid (NJDG). The NJDG is a database of orders, judgments and case details of 18,735 District & Subordinate Courts and High Courts. Data is updated on a near real-time basis by the connected district and taluka courts. It provides data relating to judicial proceedings and decisions of all computerized district and subordinate courts in India. The relevant data of all High Courts are available on the Grid [21]. Litigants may access case status, pendency data and orders of courts using the Grid. The Grid has the ability to perform drill-down analysis based on the age of the case as well as the State and District; it provides court-wise, subjective-wise and age-wise pendency of subordinate and high courts [22]. The NJDG provides the Central and State Governments with exclusive access to data using a departmental ID and access key. This facilitates institutional litigants' access to the NJDG data for monitoring and evaluation of cases where such institutions are one of the parties to pending litigations [23]. However, the Grid cannot produce predictive analyses of the case pendency and public presentation of data by NJDG is currently limited to descriptive analysis.

Despite the ongoing efforts to digitize court judgments, particularly in lower courts, many of these judgments remain in paper form. Moreover, many subordinate courts lack basic computing infrastructure, thereby causing inefficiencies in the system. The current advancements in the application of AI in India's judicial administration are overlooking the requirements of the lower courts. The lower courts are the real face of India's judiciary as they deal with 87.4% of all pending cases in the country [22]. Besides the burden of backlog, the subordinate courts face acute shortages of judicial personnel, infrastructure and funding [23]. The challenges facing the lower courts further emphasise the need for innovative and frugal solutions; AI and machine learning can address some of the challenges facing the subordinate judiciary in India by identifying the major pain points and roadblocks. It is in this background, the current study works on delay and pendency prediction with data of subordinate courts. As NJDG holds categorised data on geographical, subject-wise, court-wise, judge-wise and purpose-wise data holds great promise for predictive insights and also in discovering the relevant aspects that cause the delays. Therefore, building a predictive model addressing the delay or their causes for cases clogged in the judicial system or analysing the importance of aspects that cause the delay is a logical next step.

In this study, to explore the feasibility of predictive insights, we train a multi-class classification model using Google's Vertex AI and AutoML framework on a publicly available dataset of 4.2 million cases. The AutoML framework also identifies the feature importance matrix that we analyse for the causes of the delay. We also train a binary classification model that can more accurately predict delays for cases which take more than the average duration of delays which are found to be around 3 years in the dataset. The overall methodology followed is depicted in Fig. 1.



Process	Output	Tools
<ul style="list-style-type: none"> <li>• Data pre-processing</li> <li>• Feature Engineering</li> <li>• Multi-classification modelling</li> <li>• Binary-classification modelling</li> <li>• Model evaluation &amp; testing</li> </ul>	<ul style="list-style-type: none"> <li>• Merged CSV file</li> <li>• Feature vectors</li> <li>• Trained Classification Model</li> <li>• Trained Classification Model</li> <li>• Classification results &amp; feature importance</li> </ul>	<ul style="list-style-type: none"> <li>• Big Query (Google)</li> <li>• Custom python code</li> <li>• Vertex AI AutoML (Google)</li> <li>• Custom Python code, scikit-learn</li> <li>• Vertex AI AutoML, scikit-learn</li> </ul>

**Fig.1.** Research Methodology Workflow

## 3 Methodology

### 3.1 Dataset and Pre-processing

The dataset we use is taken from the work of Bhowmick and others [24]. The dataset was constructed from case records scraped from the e-courts platform (<https://ecourts.gov.in/>) maintained by the Indian government and includes cases from 2010 to 2018. The dataset covers data from 7,000+ district and subordinate trial courts, which are staffed by over 20,000 judges. The dataset is made available by the authors for download at <https://www.devdatalab.org/judicial-data> and contains data of over 77 million cases. Data about these cases are tracked and updated till 2020, thereby providing comprehensive coverage over the period from 2010 to 2020. There is no personally identifiable information in the dataset. The metadata descriptions for the different features are also provided by the authors.

The dataset is organized into multiple year-wise files from 2010 to 2018 in CSV and DTA format. The files include the filing dates, the first, last and next dates of hearing for the case. In cases where the judgment has been delivered the date of the decision is provided. The state, district and court details are included along with the gender details of the petitioner, defendant, and their respective lawyers. Other important data includes the position of the presiding judge, the type and the purpose of the case. Separate metadata files are also included to augment the information available in the year-wise data file. Data for the judges' gender, tenure and the applicable Act and Section for criminal offences is separately made available with a unique case Id acting as the index to the data in these files. Certain features like the date of the first, next and last listing, purpose name and disposition were dropped from the analysis as they would not be available on the filing date. Table 1 provides a summary of the features that were included in the analysis.

The overall dataset is of 77 million cases comprising cases filed from 2010 to 2018 and decided till 2020. In the current study, however, we include 4.2 million cases which are filed in the year 2010 and for which data has been captured for their outcomes till 2020. This approach ensures we include cases with the longest evaluation period from the dataset. Including cases filed from 2011 to 2018 in our model will bias the model towards cases which have been resolved early. Since the data is spread over multiple files and the files were large in size (over 1 GB) it is difficult for desktop applications like MS Excel to easily process them for data manipulation. We leveraged the Big Data warehouse Big Query from Google to merge all data into a single CSV file for the year 2010 for analysis. The total number of cases was 4.19 million.

**Table 1.** Feature Description

Column Name	Description	Usage in Model
date_of_filing	Date when the case was filed	The target variable is calculated based on the date of decision - the date of filing
date_of_decision	the case reached a decision Date when	
state_code	Unique state code	Categorical Input features for Geographical Identifiers are calculated as the combinations
dist_code	Unique District Code for every district in a state	
court_no	Unique Court number for every court in a district	
judge_position	The judge's position as per the existing hierarchy	Categorical Input feature related to the case characteristics
female_judge_filing	Gender for the judge under whom the case was filed	
female_judge_decision	Gender of the judge by whom the decision was delivered	
female_adv_pet	Gender of the advocate from the petitioner	
female_adv_def	Gender of the advocate from the defendant	
female_petitioner	Gender of the petitioner	
female_defendant	Gender of the defendant	
type_name	The type of case	
section	Section under which case is filed	
act	Act under which the case is filed	
criminal	Whether this was a criminal case	
number_sections_ipc	Number of sections charged in criminal cases	
bailable_ipc	Flag to indicate if it is bailable	

### 3.2 Feature Engineering

The transformation of target and input features was required for converting the date fields to durations and subsequently into categories. The target (or the predicted) variable in the AutoML model was defined as one of the categories of pendency for a judgement, namely “<1 year”, “1 up to 3 years”, “> 3 up to 5 years”, “>5 up to 10 years” and “> than 10 years”. The class boundaries were chosen based on the categories available in the e-courts platform where the dataset was collected. Ongoing cases, where decision dates were not available, were marked in the last category.

In the custom binary classification model the model predicted one of the two categories of “< than 3 years” and “3 years and >”. The 3-year boundary was chosen for the

binary classification model as it approximated the boundary for predicting cases with more than average delays. The average duration of delays from filing to decision for all completed cases in our dataset was found to be 36.2 months with the median value being 31 months.

Data points were dropped if there was any discrepancy in the timelines (e.g. if the filing date was older than the decision date). Cases with filing, first hearing and decision dates earlier than 2010 were also dropped from further analysis. The categories for the target variable and feature variables were obtained by calculating the intermediate period in months or days between the dates and these were used further. A few significant data variables that were dropped include judge-related data like their tenure and if during the case there was any change of the judges. The same was done for 2010 data which had a significantly large number of missing values. Missing values were imputed with a “Not Available” string token.

### 3.3 Data Split

Post-cleaning, 3.62 million (86.4% of the overall) data points remained and the standard 80:20 split for training and validation/test split was followed for model building. The AutoML model further split the validation dataset into 10% validation and 10% test data. The total number of data points available for training was 2.90.M and 0.72M for validation for the custom model. No hyper-parameter tuning was employed for the custom models and hence a three-way split as done for the AutoML model was not required. Stratifying was used during the train test split in the custom model development to ensure a proportionate representation of all target classes. Google Vertex AI, AutoML model provides out-of-the-box functionality for ensuring stratification for the train test split.

## 4 Modelling and Results

AutoML is a rapidly growing field in the domain of machine learning that aims to automate the process of building machine learning models [25]. It involves the use of algorithms and techniques to automate tasks such as feature selection, hyperparameter tuning, and model selection. Practitioners commonly use AutoML for initial prototyping to check the quality of the data and perform feature selection [26]. This usage of AutoML helps quickly identify the most relevant features and obtain a quick preliminary understanding of the relationships between the features (independent variables) and the target (dependent variable).

Once the initial validation and prototyping have been done, it is common to build custom models [27]. Decision forests and in particular, Gradient Boosting Decision Trees (GBDT) are popular choices given that they have produced state-of-the-art results in several online competitions on sites like Kaggle. The Decision forests are also easier to train and interpret [28]. Further, for developing the custom models, binarization techniques are employed where multiple binary classifiers are used to model the multi-classification problem [29]. This approach of using multiple binary classification models can provide

a more nuanced understanding of the problems. For example, if a multi-class classification model is used to predict the pendency of cases, it may not be possible to distinguish between cases that are highly likely to be resolved soon and cases that are likely to be pending for a long time. By using multiple binary models, each model can be trained to handle a different part of the prediction, allowing for a more nuanced understanding of the problem.

In our model-building process, we utilize the aforementioned methodology. First, we utilize Google's Vertex API framework to construct an AutoML multi-class classification model, which predicts pendency into five classes. Next, we use the identified features to build a custom binary classification model in Python only focusing on a binary classification of the delayed cases into greater than or less than 3 years classes. This model employs scikit-learn's decision tree and ensemble classifiers of Bagging, Random Forest, and XGBoost.

Given the need for and importance of Explainable AI (XAI) for interpretable and actionable insights in domains which impact diverse groups of people we use the SHAP [30] framework. The SHAP framework provides value plots which are a way to measure the contribution of each feature to complex prediction models. For our study, the plots can be useful in determining the most important aspects of pendency.

For the binary classification model, the feature importance matrix is readily available in the scikit-learn decision tree classifiers. However, for obtaining the feature important, Label encoders were used for the transformation of input categorical features. Label Encoders since they put an artificial order to the categorical variables are not the best-performing encoders [31], however, their use makes the feature importance matrix readily available. The interpretation of the feature importance matrix for the models facilitates the prioritization of areas where the focus needs to be provided for improving the court processes for pendency reduction.

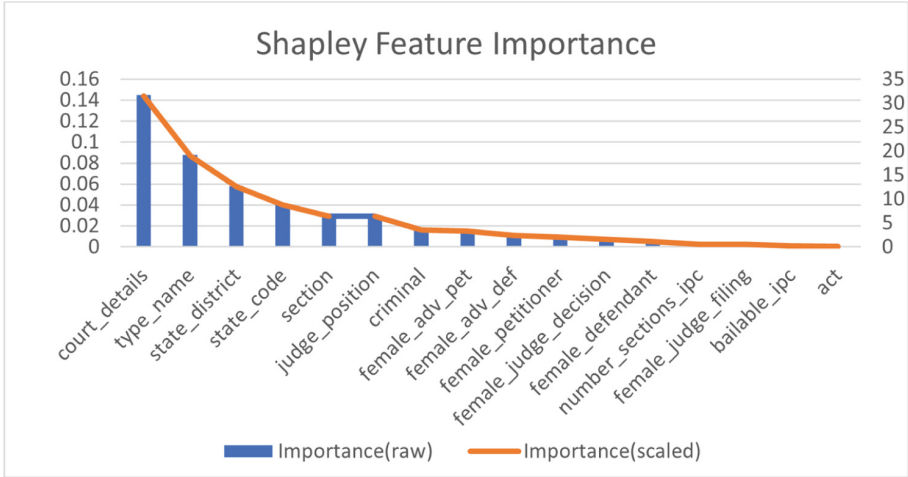
#### 4.1 AutoML Model

**Model Training.** Google AutoML Vertex AI automates the process of finding the best model and tuning its hyperparameters to achieve the best performance. The training time was 9 h and 20 min which included 300 tuning trials with different Neural Networks and Boosted Tree Models with different configurations of layers and trees respectively. The best-performing model which was proposed by AutoML was an ensemble of 25 neural network models with 2, 3 or 5 hidden layers with sizes of 16, 128 or 1024. Dropout values used in the different models were 0.25, 0.375 or 0.5.

**Results.** The results showed varying levels of accuracy for the different classes. The model was best performing in "< than 1 year" and "5 to 10 years" with an accuracy of 83% and 72% respectively, with F1 scores of 0.66 and 0.54, PR AUC of 0.77 and 0.66, ROC AUC of 0.90 and 0.85, Precision of 73.2% and 69.7% and Recall of 60.7% and 43.8%. Overall, the model had an F1 score of 0.47, PR AUC of 0.60, ROC AUC of 0.86, Precision of 62.9% and Recall of 37.3%.

The SHAP feature importance metrics indicated that the geographic identifiers namely the specific courts, specific districts and the state contributed almost 50% while the type of case, section and the judge position were the other important features. The

detailed results are presented further in Fig. 2 (feature importance), Table 2 (confusion metrics) and Table 3 (accuracy metrics). The findings indicated that the model performance was good in classifying cases only in <1 year and 5 to 10 years band, but the performance was found inadequate in all other classes. The feature importance metrics indicated that the shortlisted features (16 in number) all contributed to the model prediction.



**Fig. 2.** Feature Importance (AutoML)

**Table 2.** Confusion Matrix (AutoML)

True /Predicted label	<1 year	1 upto 3 years	3 upto 5 years	5 upto 10 years	>10 years
<1 year	<b>83%</b>	16%	0%	0%	0%
1 to 3 years	27%	<b>55%</b>	13%	6%	0%
3 to 5 years	33%	8%	<b>27%</b>	33%	0%
5 to 10 years	10%	3%	16%	<b>72%</b>	0%
>10 years	5%	0%	2%	93%	0%



**Table 3.** Accuracy Matrix (AutoML)

	All Labels	<1 Year	1 upto 3 Years	>3 upto 5 Years	>5 upto 10 Years	> 10 years
PR AUC	0.60	0.76	0.51	0.36	0.66	0.02
ROC AUC	0.86	0.90	0.72	0.74	0.85	0.58
Log loss	0.99	0.34	0.55	0.47	0.38	0.001
F1 score	0.47	0.66	0.39	0.19	0.54	0.07
Precision	62.9%	73.2%	54.7%	37.6%	69.7%	100%
Recall	37.3%	60.7%	29.8%	12.4%	43.8%	4.5%

## 4.2 Custom Model

**Model Training.** The custom model was developed in Python using the Google Colab Pro development environment. The scikit-learn, Python and machine learning packages were used for the modelling. Google Colab Pro was used as the RAM requirements were larger than those provided by the free version of Google Colab and consequently lead to code crashes. The code file, along with the cleaned dataset is made available at github at <https://github.com/mb7419/pendencyprediction>, for easy reproducibility of the results. The custom model included all the features included in the AutoML model since all showed some contributions in the feature importance. As discussed earlier the target variable was collapsed into two classes, one with “< than 3 years” and the other one was “3 years and >”.

The decision tree classifier was used as a baseline, with a max tree depth of 10. The tree depth was set to avoid overtraining. Further the ensemble classifiers of Random Forest, Bagging and XGBoost were used to improve on the baseline model. The binary classification metrics, equivalent to the AutoML model, were used for evaluation. Label encoder was used for transforming the categorical features in these models as it is difficult to interpret when one-hot-encoding is used.

The feature importance matrix was determined based on the classifier that gave the best-performing model. The classifier that gave the highest accuracy and performance was also used to train a separate model using scikit-learns one-hot-encoding. TruncatedSVD was further used for dimensionality reduction with the top 200 components used for modelling. This model was developed to get improved or equivalent performance to the label encoder-based categorical feature encoding.

**Results.** Random Forest achieved the best accuracy scores among the four different classifiers used. An accuracy of 81.5%, and the weighted average precision, recall and F1 were found to be 0.82, PR AUC was 0.40 and ROC AUC of 0.81. The confusion matrix for the results is presented in Fig. 3. The model built by using One-hot-encoding of the categorical variables improved the accuracy to 81.7% and the performance of the model was almost similar. The Bagging classifier was a close second in classification performance and the XGBoost out-of-the-box also provided similar performance. We have not conducted further hyperparameter tuning as the performance of the Random

Forest classifier was found adequate for our use case and validated the hypothesis that it was possible to improve the performance of the AutoML model with a custom binary classification model. The comparative result with all classifiers is available in Table 4 and the Feature Importance is available in Fig. 4. The overall results indicate that Random Forest with high performance across all metrics was a robust classifier for predicting pendency in our dataset.

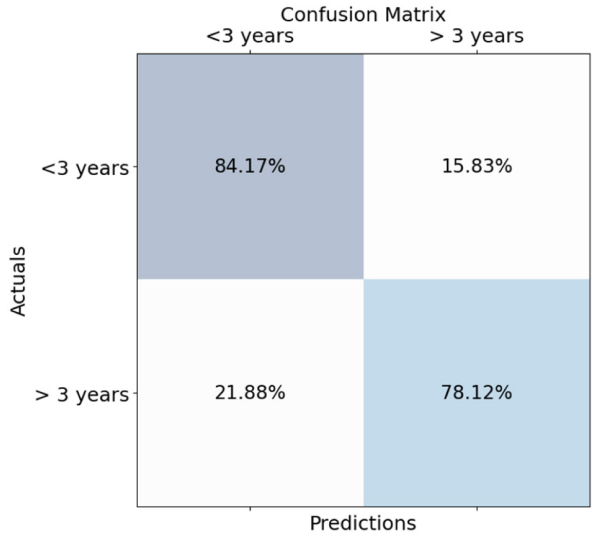


Fig. 3. Confusion Matrix – Custom

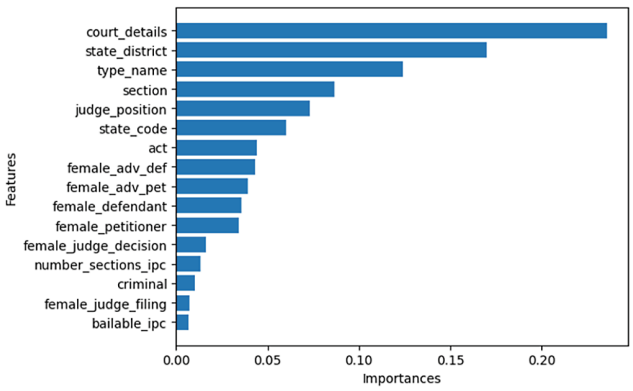


Fig. 4. Feature Importance (Custom Model)

**Table 4.** Comparative Results (Custom Model)

	Decision Tree	Bagging	XGBoost	Random Forest
Accuracy	0.74	0.81	0.81	0.82
PR AUC	0.33	0.39	0.39	0.40
ROC AUC	0.72	0.81	0.80	0.81
Log loss	0.52	0.80	0.41	0.57
F1 score	0.73	0.81	0.81	0.82
Precision	0.74	0.81	0.81	0.82
Recall	0.74	0.81	0.81	0.82

## 5 Discussion of Key Results

The results demonstrate the feasibility of predicting delays in lower courts using available data during the early stages of the cases. The AutoML results indicate a high level of accuracy in predicting both extreme cases, specifically those with durations of less than one year and greater than ten years. Furthermore, employing a random forest binary classification model yielded comparable and satisfactory values for accuracy, F1 score, precision, and recall metrics. This model exhibited robust performance in predicting cases with delays ranging from less than three years to more than three years. The modelling revealed a feature matrix, which highlights the key factors that contribute to delays in court proceedings. Notably, the type of court, characteristics of the case, and geographical location were found to have a significant influence on the occurrence of delays (see Fig. 4). Additionally, the subject matter, such as the specific Act involved, and the complexity of the criminal code sections were also identified as factors contributing to delays. Moreover, the participation of women in the justice administration, including roles as judges, advocates, or litigants, displayed a positive correlation with case delays (see Fig. 4). These findings align with and support existing literature on the subject, except on the correlation between the participation of women in justice administration and the delay that needs further exploration.

### 5.1 State, District, Court Type and Judge Position

India has a hierarchical but unified judiciary. The judiciary is arranged into three tiers, namely the Union Judiciary (the Supreme Court), the State Judiciary (High Courts), and Subordinate Courts [31]. Administration justice at the high court level and below is supported primarily by the state governments. Similarly, the population of states, territorial limits of courts, and their operational efficiency vary significantly from one state to another [32]. There are demographic and infrastructural divergences that have a bearing on the pendency. The volume, average length of pendency and case clearance rate varies significantly among different states [33]. Even within a state, there are significant divergences in pendency [34]. Likewise, there is notable divergence in the volume of the type of cases (civil and criminal) among the states [34]. A strong correlation between

state, court, and judge types and pendency, as shown in Fig. 2, is consistent with the literature [33, 35].

## 5.2 Act Type, Case Types (Criminal), Number of Sections and Bailable Offences

Compared to petty (bailable) criminal cases, summons criminal cases take a longer time to reach finality. 90.09% of the pending criminal cases are summons cases [33]. Additionally, the number of charges under the Indian Penal Code, case and Act types are positively correlated with the length of the pendency. The stages at which cases experience delays vary, with the severity of charges and the number of sections involved playing a role in the length of pendency. For example, according to the Daksha Report, 31% of cases were pending at the “appearance” stage, 7.2% at the “framing of charges” stage, 36% at the “evidence” stage, 17.9% at the “arguments” stage, and 8% at the orders stage [34].

The severity of charges in non-bailable offences suggests that they tend to remain pending in the judicial system for a longer duration compared to bailable offences. Delays are more pronounced in session cases, which involve serious criminal offences, compared to summons cases. Furthermore, the number of charges under the Indian Penal Code, as well as the type of case and Act, positively correlates with the length of pendency. Therefore, the case-clearing rate is influenced by factors such as the type of Act, case types, the number of sections involved, and bailable offences. The modelling rightly underlines the correlation between these features.

## 5.3 Role of Gender in Pendency

In India, the overall numbers of women judges in lower courts remain low; however, in recent decades, higher proportions of women have been appointed to the lower judiciary [36]. Despite the presence of a considerable number of women judges in the judiciary, little empirical inquiry has been made into whether their experiences of judging and judicial work differ and whether they face distinct challenges that affect their optimal performance [37]. This study points out that there might be some considerable occupational challenges that women judge face that hinder their performance. Among the features that impact the duration or outcome of legal cases, six relate to the participation of women in various roles. Out of these six features, two were related to female judges. This study also indicates that the judicial administration may not be adequately responsive to and accommodating of women, as the findings reveal that court cases are more likely to experience delays when one of the parties is female, when female lawyers represent the parties, or when a female judge presides over the case. These results are somewhat unexpected, but not surprising as the gender (decisional) bias and inadequate representation of women at the bar and on the bench are well established [38–40]. However, the findings of this research call into question the equal application of procedural law and court processes. There is a need for further empirical investigation to test the findings of this study.

## 6 Limitations and Future work

The current modelling of pendency in Indian courts incorporates some factors that are known to contribute to delays, but there is significant room for improvement by bringing in more factors like the judge-specific data. The processing of some high cardinal categorical variables used label encoders & one-hot-encoding for encoding, the processing of which can be improved by using Hashing Encoder [41, 42]. The model can be made more robust by including data from 2011 onwards which is available in the dataset and further fine-tuning of the models can be implemented for improving accuracy and performance.

A predictive application can also be developed which can provide this information readily to the various stakeholders involved. The absence of data on certain features such as interlocutory applications, inter-court appeals, and stays of proceedings by higher courts limits the ability to examine their impact on case duration. To address this, a more comprehensive data set can be further constructed to increase the accuracy and insights of the predictive model. In particular, more detailed data on the categorisation at the court and district levels is required to forecast pendency more effectively. Although the existing data set provides information on courts, districts and states, unique codes are needed to locate specific courts or districts within a state, which would improve the ability to predict the duration of the case.

## References

1. Law Commission of India: Reform of Judicial Administration (Law Comm No 14, 1958) para 4.2; Over 71,000 Cases Pending in Supreme Court, 59 Lakhs in High Courts: Law Minister Tells Rajya Sabha. Live Law (2022). <https://www.livelaw.in/top-stories/over-71000-cases-pending-in-supreme-court-59-lakhs-in-high-courts-law-minister-tells-rajya-sabha-205784>. Accessed 11 July 2023
2. Verma, K.: E-courts project: a giant leap by Indian judiciary. *Legal Inf. Inst.* **4**(12), 1–12 (2018)
3. Khaitan, N.: Inefficiency and Delay. *Vidhi Cent. Legal Policy* 1–24 (2017)
4. E-Committee: Supreme Court of India. National Policy and Action Plan for Implementation of Information and Communication Technology in the Indian Judiciary (2005). <https://main.sci.gov.in/pdf/ecommittee/action-plan-ecourt.pdf>
5. National Judicial Data Grid Homepage. <https://njdg.ecourts.gov.in/njdgnew/index.php>
6. 50 million cases were still unresolved as of 2022, including more than 169,000 cases in district and high courts that had been languishing for more than 30 years. Over 85% of all cases—43 million out of 50 million—were still pending in district courts as of December 2022
7. Cui, Y.: Artificial Intelligence and Judicial Modernization. Shanghai People's Publishing House (2020). <https://doi.org/10.1007/978-981-32-9880-4>
8. Kinhal, D., Jauhar, A., et al.: Virtual courts in India: a strategy paper. Vidhi Centre for Legal Policy (2020). [https://vidhilegalpolicy.in/wp-content/uploads/2020/07/20200501\\_\\_Strategy-Paper-for-Virtual-Courts-in-India\\_Vidhi-1.pdf](https://vidhilegalpolicy.in/wp-content/uploads/2020/07/20200501__Strategy-Paper-for-Virtual-Courts-in-India_Vidhi-1.pdf)
9. Chau, K.W.: Prediction of construction litigation outcome – a case-based reasoning approach. In: Ali, M., Dapoigny, R. (eds.) *IEA/AIE 2006. LNCS (LNAI)*, vol. 4031, pp. 548–553. Springer, Heidelberg (2006). [https://doi.org/10.1007/11779568\\_59](https://doi.org/10.1007/11779568_59)
10. Mahfouz, T., Kandil, A.: Litigation outcome prediction of differing site condition disputes through machine learning models. *J. Comput. Civ. Eng.* **26**, 298–308 (2012)

11. Aletras, N., et al.: Predicting judicial decisions of the European Court of Human Rights: a natural language processing perspective. *PeerJ Comput. Sci.* **2**, 1–19 (2016). <https://discovery.ucl.ac.uk/id/eprint/1522370/1/peerj-cs-93.pdf>
12. Şulea, O.-M., Zampieri, M., Vela, M., van Genabith, J.: Predicting the law area and decisions of French supreme court cases. In: *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017, Varna, Bulgaria*, pp. 716–722. INCOMA Ltd. (2017)
13. Strickson, B., De La Iglesia, B.: Legal judgement prediction for UK courts. In: *Proceedings of the 2020 3rd International Conference on Information Science and System*, pp. 204–209 (2020)
14. Lage-Freitas, A., Allende-Cid, H., Santana, O., de Oliveira-Lage, L.: Predicting Brazilian court decisions. *PeerJ Comput. Sci.* **8** (2022)
15. Shaikh, R.A., Sahu, T.P., Anand, V.: Predicting outcomes of legal cases based on legal factors using classifiers. *Procedia Comput. Sci.* **167**, 2393–2402 (2020)
16. Medvedeva, M., et al.: Rethinking the field of automatic prediction of court decisions. *Artif. Intell. Law* **31**, 195–212, 198 (2023)
17. Medvedeva, M., Vols, M., Wieling, M.: Using machine learning to predict decisions of the European court of human rights. *Artif. Intell. Law* **28**, 237–266 (2020)
18. Supreme Court Vidhik Anuvaad Software' is a machine-assisted translation tool trained by AI. Supreme Court of India. Press Release (2019). <https://main.sci.gov.in/pdf/Press/press%20release%20for%20law%20day%20celebratoin.pdf>
19. SUPACE: Supreme Court Portal for Assistance in Courts Efficiency; Shanti, S.: Behind SUPACE: The AI Portal of the Supreme Court of India (2021). <https://analyticsindiamag.com/behind-supace-the-ai-portal-of-the-supreme-court-of-india/>
20. Koshy, J.: Can Artificial Intelligence, Machine Learning put judiciary on the fast track? *The Hindu*, New Delhi (2022). <https://www.thehindu.com/news/national/ai-ml-are-a-long-way-from-becoming-a-judicial-decision-making-tool/article65193656.ece>
21. Department of Justice. National Judicial Data Grid (2023). <https://doj.gov.in/the-national-judicial-data-grid-njdg/>
22. Sumeda: The clogged state of the Indian judiciary. *The Hindu* (2022). <https://www.thehindu.com/news/national/indian-judiciary-pendency-data-courts-statistics-explain-judges-ramana-chief-justiceundertrials/article65378182.ece>
23. Supreme Court of India: Subordinate Courts of India: A Report on Access to Justice 2016 (Centre for Research & Planning, New Delhi), pp. 4–12 (2016)
24. Bhowmick, A., et al.: In-group bias in the Indian judiciary: evidence from 5.5 million criminal cases. In: *ACM SIGCAS Conference on Computing and Sustainable Societies*, pp. 47–47 (2021)
25. He, X., Zhao, K., Chu, X.: AutoML: a survey of the state-of-the-art. *Knowledge-Based Systems* **212** (2021). <https://doi.org/10.1016/j.knsys.2020.106622>
26. Xin, D., Wu, E.Y., Lee, D.J.L., Salehi, N., Parameswaran, A.: Whither AutoML? Understanding the role of automation in machine learning workflows. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–16 (2021). <https://doi.org/10.1145/3411764.3445306>
27. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems*, vol. 30 (2017)
28. Sagi, O., Rokach, L.: Approximating XGBoost with an interpretable decision tree. *Inf. Sci.* **572**, 522–542 (2021). <https://doi.org/10.1016/j.ins.2021.05.055>
29. Galar, M., Fernández, A., Barrenechea, E., Bustince, H., Herrera, F.: An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recogn.* **44**(8), 1761–1776 (2011). <https://doi.org/10.1016/j.patcog.2011.01.017>

30. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems*, Long Beach, CA, USA, pp. 4765–4774 (2017)
31. The Constitution of India 1950, Pt VI, Ch 4 and 5
32. Share of case clearance rate in subordinate courts across India in the financial year 2019, by State. Statista (2019). <https://www.statista.com/statistics/1211859/india-share-of-case-clearance-rate-in-subordinate-court-by-state/>
33. India Justice Report 2022, Sir Dorabji Tata Trust, 89–92, 97–99 (2023). [https://indiajustice-report.org/files/IJR%202022\\_Full\\_Report1.pdf](https://indiajustice-report.org/files/IJR%202022_Full_Report1.pdf)
34. Kaul, A., et al.: *Deconstructing Delay: Analyses of Data from High Courts and Subordinate Courts* (2023). [https://www.dakshindia.org/Daksh\\_Justice\\_in\\_India/19\\_chapter\\_01.xhtml#\\_idTextAnchor094](https://www.dakshindia.org/Daksh_Justice_in_India/19_chapter_01.xhtml#_idTextAnchor094)
35. Pratik Dutta and Suyash Rai: How to Start Resolving the Indian Judiciary’s Long-Running Case Backlog (2021). <https://carnegieendowment.org/2021/09/09/how-to-start-resolving-indian-judiciary-s-long-running-case-backlog-pub-85296>>; National Judicial Data Grid: Evidence/Argument/Judgement Wise Pendency, [https://njdg.ecourts.gov.in/njdgnew/?p=main/pend\\_dashboard](https://njdg.ecourts.gov.in/njdgnew/?p=main/pend_dashboard)
36. The national average of women’s representation in the lower judiciary is 35%, see India Justice Report 2022, 15, 94 (2023)
37. Tannvi, T., Narayana, S.: The challenge of gender stereotyping in Indian courts. *Cogent Soc. Sci.* **8**, 1 (2022)
38. Pragati K.B.: *Skewed Corridors of Justice: Women Continue to Face Sexism in Courts*. The Wire (2019). <https://thewire.in/women/sexism-courts-women-lawyers-judges>
39. Vidhi Centre for Law & Policy Search. *Tilting the Scale* (2018). [https://vidhilegalpolicy.in/wp-content/uploads/2020/06/180212\\_TiltingtheScale\\_Final.pdf](https://vidhilegalpolicy.in/wp-content/uploads/2020/06/180212_TiltingtheScale_Final.pdf)
40. Satish, M.: *Discretion. Discrimination and the Rule of Law: Reforming Rape Sentencing in India*. Cambridge University Press; Gender Diversity Portal, Vidhi Centre for Legal Policy (2016). [https://data.vidhilegalpolicy.in/dashboard/gender\\_diversity/gender\\_diversity.htm](https://data.vidhilegalpolicy.in/dashboard/gender_diversity/gender_diversity.htm)
41. Cerda, P., Varoquaux, G.: Encoding high-cardinality string categorical variables. *IEEE Trans. Knowl. Data Eng.* **34**(3), 1164–1176 (2020)
42. Hancock, J.T., Khoshgoftaar, T.M.: Survey on categorical data for neural networks. *J. Big Data* **7**(1), 1–41 (2020)



# Chaotic Image Encryption Using an Improved Vigenère Cipher and a Crossover Operator

Mourad Kattass<sup>1</sup>(✉), Hicham Rrghout<sup>1</sup>, Mariem Jarjar<sup>1</sup>, Abdellatif Jarjar<sup>1</sup>,  
Faiq Gmira<sup>2</sup>, and Abdelhamid Benazzi<sup>1</sup>

<sup>1</sup> MATSI Laboratory, Mohammed First University, Oujda, Morocco  
{Mourad.kattass,h.rrghout,a.benazzi}@ump.ac.ma

<sup>2</sup> Innovative Technologies Laboratory (LTI), University Sidi Mohamed Ben Abdellah, Fez,  
Morocco

faiq.gmira@usmba.ac.ma

**Abstract.** In this article, we will explore a new technique for encrypting color images. This approach is based on the classical Vigenère technique which has been improved and adapted to image encryption. This approach involves the use of two substitution tables of the same size, generated from the chaotic maps. Then, a crossover operation is applied on the integrity of the vector leaving the substitution phase, this operation is controlled by a pseudo-random vector.

The standard security tests applied to our technique proves the robustness and effectiveness of our approach against known attacks.

**Keywords:** image encryption · S-Box · new Vigenère's replacement functions · Chaotic map · Chaotic genetic cross-over · encryption function

## 1 Introduction

With the increasing importance of digital security and privacy, encryption techniques are becoming more important than ever [1–3]. In this context, researchers are constantly looking for new and innovative ways to protect sensitive information from unauthorized access[1–3]. One such technique is the encryption of color images, which is the focus of this article.

The approach discussed in this article is based on the classical Vigenère technique [4–6], which has been around for centuries. However, it has been improved and adapted to image encryption. This new approach involves the use of two substitution tables of the same size, which are generated from chaotic maps [7, 8]. The use of chaotic maps adds entropy to the encryption process, making it harder for attackers to break it.

To verify the robustness and effectiveness of this new approach, the standard security tests were applied. These tests have proven that our technique is highly robust and effective against known attacks, making it a reliable option for safeguarding sensitive color images.

Once these substitution tables are created, a crossover operation [9, 10] is applied to the integrity of the vector leaving the substitution phase. This operation is controlled by



a pseudo-random vector, which adds another layer of security to the encryption process. This ensures that the final encrypted image is highly secure and difficult to decrypt without the appropriate key.

In previous work, we identified the following publications:

In reference [10] the authors suggested a new image encryption technique based on the Vigenère cipher using an S-box of size 16x16, by integrating genetic mutations.

The authors of the article [11] have developed a method for encrypting color images using two Vigenère tours separated by a mutation operation.

In the article [12] A. Abdellah et al. proposed a cryptosystem using a single Vigenère tour followed by a chaotic permutation on the integrity of the images.

## 2 The Proposed Method

The method we utilize involves utilizing several generations of chaotic sequences. To accomplish this, we rely on the utilization of two specific chaotic maps, namely the logistics map ( $L_n$ ) and the skew tent map ( $S_n$ ).

### 2.1 The Logistics Map ( $L_n$ )

This is a 1D chaotic sequence, and it's one of the most commonly used chaotic sequences in several applications. The logistic sequence was popularized by biologist Robert May in 1976. It's defined by the following simple nonlinear recurrence [8]:

$$\begin{cases} L_0 \in ]0, 1[ \\ L_{n+1} = \mu L_n (1 - L_n) \end{cases} \quad (1)$$

$L_0$  and  $\mu$  are control parameters. The chaotic behavior of the logistic sequence is obtained if  $\mu \in [3.57, 4]$  and  $L_0 \in ]0.5, 1[$ .

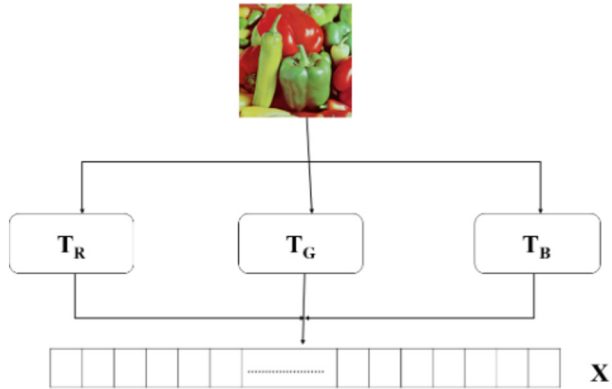
### 2.2 The Skew Tent Map ( $S_n$ )

The skew Tent map is a nonlinear dynamic system with complex chaotic behavior. It is one of the most used chaotic maps, it is expressed by the following equation [8]:

$$\begin{cases} S_0 \in ]0, 1[ \\ S_{n+1} = \begin{cases} \frac{z_n}{\delta} & \text{if } 0 < S_n < \delta \\ \frac{1-S_n}{1-\delta} & \text{if } \delta < S_n \end{cases} \end{cases} \quad (2)$$

### 2.3 Preparation of the Original Image

The process begins by loading the original image of size (N, M). This image is then divided into three distinct color channels (Red, Green and Blue). The matrices R, G and B are transformed into three vectors  $T_R$ ,  $T_G$  and  $T_B$  of size  $1 \times NM$ . These three vectors are then concatenated to form a single vector X of size  $1 \times 3NM$ .



**Fig. 1.** The preparation of the original image

Figure 1 illustrates this process.

This transformation is illustrated by the following algorithm:

---

**ALGORITHM 1:** Convert to vector

---

```

For j=0 to NM-1
    X(3j) = TR(j)
    X(3j+1) = TG(j)
    X(3j+2) = TB(j)
Next j
    
```

---

**2.4 Design of Encryption Keys**

Our work is based on the construction of five chaotic vectors AR1, AR2, AR3, AR4 and AR5 in  $Z/256Z$  and two binary control vectors CA and BA. This construction is illustrated by Algorithm 2.

---

**ALGORITHM 2: Generation of pseudo random vectors**


---

```

For  $i=0$  to  $3nm-1$ 
  AR1 ( $i$ ) =  $\text{mod}(E(|l(i) - s(i)| * 10^{11}, 253) + 3)$ 
  AR2 ( $i$ ) =  $\text{mod}(E(l(i) * s(i) * 10^{12}, 253) + 2)$ 
  AR3 ( $i$ ) =  $\text{mod}(E(l(i) * 10^{11}, 253) + 1)$ 
  AR4 ( $i$ ) =  $\text{mod}(E(s(i) * 10^{10}, 253) + 1)$ 
  AR5 ( $i$ ) =  $\text{mod}(\max(\text{AR1}(i), \text{AR4}(i)), 5) + 1$ 
  if  $l(i) \geq s(i)$ 
    then CA ( $i$ ) = 0 else CA ( $i$ ) = 1
  endif
  if AR1 ( $i$ ) > AR3 ( $i$ )
    then BA ( $i$ ) = 0 else BA ( $i$ ) = 1
  endif
Next  $i$ 

```

---

## 2.5 Advanced Vigenère Method

This technique requires the establishment of two substitution matrices ( $TB1$ ) and ( $TB2$ ) of size (256, 256), through the process described by the following steps

- a permutation ( $Q1$ ) obtained by descending ordering the first 256 values of the sequence ( $AR1$ )
- a permutation ( $Q2$ ) obtained by descending ordering the first 256 values of the sequence ( $AR2$ ),

The expansion of the two substitution tables is given by the following algorithm:

---

**ALGORITHM 3: S-Box design**

---

```

//S-Box design
//Fist Row
For i=0 to 255
    TB1(0,i)=Q1(i)
    TB2(0,i)=Q2(i)
Next i
//Next lines
For i=1 to 255
    For j=0 to 255
        if BA(i)=1 then
            TB1(i,j)=TB1(i-1,j $\oplus$ AR3(i))
            TB2(i,j)=TB2(i-1,j $\oplus$ AR4(i))
        else
            TB1(i,j)=TB1(i-1,j $\oplus$ AR1(i))
            TB2(i,j)=TB2(i-1,j $\oplus$ AR2(i))
        endif
    Next j,i

```

---

**2.6 Initialization Value Design.**

First, the (Iv) initialization value must be recalculated to change the value of the starting pixel. Ultimately, the (Iv) value is provided by the next algorithm:

---

**ALGORITHM 4: Initialization value design**

---

```

Iv=0
For i=0 to 3NM-1
    if BA(i)==0
        Iv=X(i) $\oplus$ Iv $\oplus$ AR1(i)
    else
        Iv=X(i) $\oplus$ Iv $\oplus$ AR2(i)
    endif
Next i
X(0)=Iv  $\oplus$  X(0)
Y(0)=TB1(AR3(0), TB2(AR1(i); (mod(a*X(0), 256) $\oplus$ AR3(0)))

```

---

The value calculated from the clear image and the chaotic map, will only be used to change the value of the start pixel and restart the encryption process. The new confusion function and new diffusion function are given by the following algorithm:

ALGORITHM 5: Encryption process

```

For i=1 to (3NM-1)
#diffusion
  X(i)= Y(i-1) ⊕ X(i)
#Confusion
if CA(i)=0 then
  Y(i)=TB1 (AR3 (i) , TB2 (AR1 (i) ; (mod (a*X (i) , 256) ⊕ AR3 (i) ))
  else
  Y(i)=TB2 (AR1 (i) , TB1 ( (AR2 (i) ; (mod (c*X (i) , 256) ⊕ AR4 (i) )))
Next i
    
```

With:  $a = \text{mod}(2 * \text{mod}(\sum_{i=1}^n AR1(i)) + 3,256)$   
 $c = \text{mod}(2 * \text{mod}(\sum_{i=1}^{nm} AR2(i)) + 1,256)$

2.7 Diagram of the Encryption Process

(See Fig. 2).

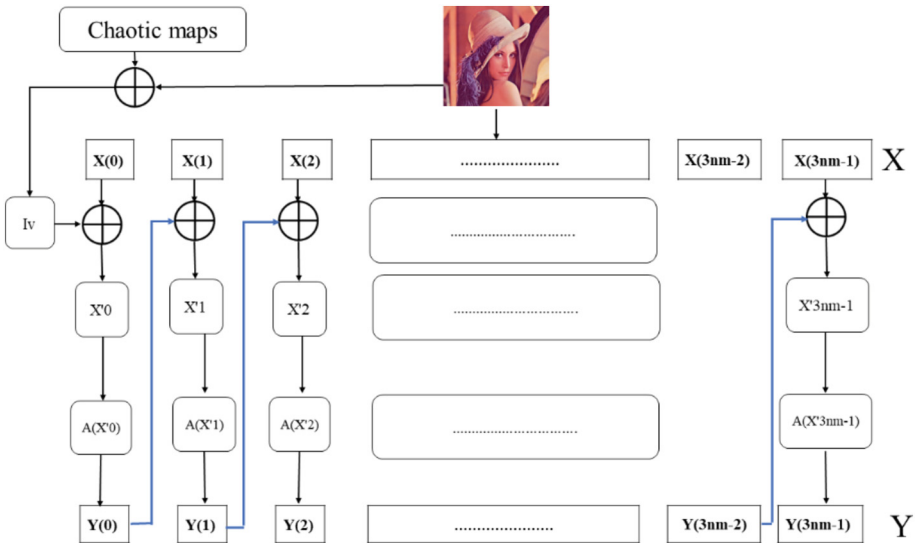


Fig. 2. First round

2.8 The Genetic Crosse-Over

A genetic crossover is the interaction between two parent genes to reproduce a new daughter gene which inherits the characteristics of the parents. Such a crossing will be under the control of the vector (AR5). The crossing is illustrated by the algorithm below:

**ALGORITHM 6: Genetic Crosse-over**


---

```

for i=0 to 3nm-1
If AR5(i)=1 Then Z(i)=Y(i) ⊕ AR1(i)
If AR5(i)=2 Then Z(i)=Y(i) ⊕ AR2(i)
If AR5(i)=3 Then Z(i)=Y(i) ⊕ AR3(i)
If AR5(i)=4 Then Z(i)=Y(i) ⊕ AR4(i)
If AR5(i)=5 Then Z(i)=Y(i) ⊕ AR5(i)
Next i

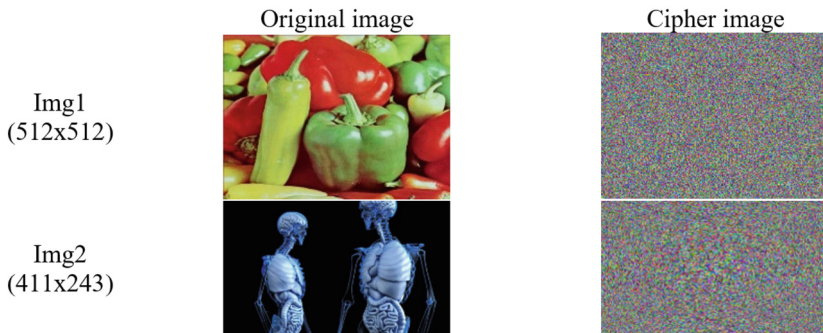
```

---

The output vector ( $Z$ ) represents the image encrypted by our new technology.

### 3 Simulation and Analysis

Our experiments were conducted on a personal computer running Windows 10 and utilizing the Python programming language. The computer's configuration includes an Intel(R) Core(TM) i7-6600U @ 2.60 GHz processor 2.80 GHz, 8 GB of RAM, and a 512 GB SSD hard drive. The figure below illustrates the images encrypted by our method (Fig. 3).



**Fig. 3.** The original and encrypted images

#### 3.1 Key-Space Analysis

To guarantee robustness against brute force attacks, the key space needs to encompass all possible secret key options. A wide enough key space is essential, as demonstrated by cryptosystems with key-spaces larger than 2100, which are currently impervious to brute force attacks. Our total key size exceeds  $2^{100}$  ( $2^{128} > 2^{100}$ ), ensuring sufficient protection against any potential brute force attempts.

### 3.2 Statistic Attacks

#### 3.2.1 Histogram Analysis

The histogram is a visualization tool that represents the distribution of pixel intensity values in an image. It is widely used in image encryption.

The following figure displays the histograms of the encrypted images alongside those of the original images (Fig. 4).

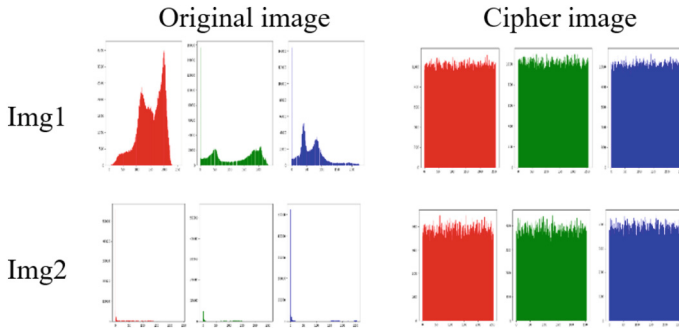


Fig. 4. Encrypted image histogram

One can observe that the uniform distribution of the histograms in the encrypted images compared to the original images guarantees the protection of our method against any possible histogram attack.

#### 3.2.2 Entropy Analysis

The entropy of an image of size (n, m) is given by the following equation:

$$H(MC) = \sum_{i=1}^{256} -pr(i) \log_2(pr(i)) \tag{3}$$

pr(i) is the probability of occurrence of level (i) in the original image. The entropy values of two images encrypted by our system are presented in Table 1 below.

Table 1. Entropy Analysis

	Entropy
Img1	7.99971
Img2	7.99952

Note that the entropy of all the images tested by our method is close to 8.

### 3.2.3 Correlation Analysis

The correlation of an image of size (n, m) is given by the equation below:

$$Corr = \frac{cov(x, y)}{\sqrt{Var(x)}\sqrt{Var(y)}} \tag{4}$$

**Table 2.** Correlation analysis

	Original image			Cipher image		
	H	V	D	H	V	D
Img1	0.683	0.476	0.517	0.0019	-0.0003	-0.0013
Img2	0.971	0.851	0.915	0.0002	0.0034	0.0037

Table 2 indicates that the correlation measurements for the images analyzed by our system are extremely close to zero, demonstrating the robustness of our approach against correlation attacks.

### 3.2.4 Differential Analysis

The parameters employed to handle differential attacks in cryptography include the Number of Pixels Change Rate (NPCR), unified average changing intensity (UACI), and avalanche effect (AE).

The NPCR is determined by the equation below:

$$NPCR = \left( \frac{1}{nm} \sum_{i,j=1}^{nm} M(i, j) \right) * 100 \tag{5}$$

$$With M(i, j) = \begin{cases} 1 & \text{if } C_1(i, j) \neq C_2(i, j) \\ 0 & \text{if } C_1(i, j) = C_2(i, j) \end{cases}$$

The UACI of an image is given by the following equation:

$$UACI = \left( \frac{1}{nm} \sum_{i,j=1}^{nm} AbsoluteValeur(C_1(i, j) - C_2(i, j)) \right) * 100 \tag{6}$$

The AE is given by the following equation:

$$AE = \left( \frac{The\ sum\ of\ bits\ changed}{The\ sum\ of\ the\ bits} \right) * 100 \tag{7}$$

Table 3 below presents the NPCR, UACI and EA values of the images encrypted by our technique.

Our method exhibits resistance to differential attacks as the differential parameters adhere to international standards, such as NPCR surpassing 199.60%, UACI outperforming 33.34%, and EA exceeding 50%.



**Table 3.** Differential parameters

Image	NPCR	UACI	AE
Img1	100%	33.609820%	50.2173%
Img2	99.9998%	33.6333%	50.2109%

## 4 Conclusion

The great sensitivity to beginning circumstances in the realm of cryptography has led to the rise of chaos theory. This new technique, which is based on chaos, uses a single Vigenère circuit to construct two brand-new, enormous S-Boxes that were created from the chaotic boards that were utilized. The new replacement functions for additional broadcast functions are closely related to these two substitution tables. A genetic crisis that was adapted to the encryption of color images was born at the conclusion of this cycle, and the findings were positive. Further studies can generalize these findings.

As future perspectives: A crossing acting at the DNA level.


## References

1. Kumari, M., Gupta, S., Sardana, P.: A survey of image encryption algorithms. *3D Res.* **8** (2017). <https://doi.org/10.1007/s13319-017-0148-5>
2. Zia, U., et al.: Survey on image encryption techniques using chaotic maps in spatial, transform and spatiotemporal domains. *Int. J. Inf. Secur.* **21**, 917–935 (2022). <https://doi.org/10.1007/s10207-022-00588-5>
3. Fang, P., Liu, H., Wu, C., Liu, M.: A survey of image encryption algorithms based on chaotic system. *Vis. Comput.* **39**, 1975–2003 (2023). <https://doi.org/10.1007/s00371-022-02459-5>
4. Li, S., Zhao, Y., Qu, B., Wang, J.: Image scrambling based on chaotic sequences and Vigenère cipher. *Multimed Tools Appl.* **66**, 573–588 (2013). <https://doi.org/10.1007/s11042-012-1281-z>
5. Qobbi, Y., Jarjar, A., Essaid, M., Benazzi, A.: Image encryption algorithm using dynamic permutation and large chaotic S-box. *Multimed. Tools Appl.* (2022). <https://doi.org/10.1007/s11042-022-14175-2>
6. Zhang, Y., Xiao, D., Wen, W., Nan, H.: Cryptanalysis of image scrambling based on chaotic sequences and Vigenère cipher. *Nonlinear Dyn.* **78**, 235–240 (2014). <https://doi.org/10.1007/s11071-014-1435-9>
7. Ali, T.S., Ali, R.: A new chaos based color image encryption algorithm using permutation substitution and Boolean operation. *Multimed. Tools Appl.* **79**(27–28), 19853–19873 (2020). <https://doi.org/10.1007/s11042-020-08850-5>
8. Zhou, Y., Bao, L., Chen, C.L.P.: A new 1D chaotic system for image encryption. *Signal Process.* **97**, 172–182 (2014). <https://doi.org/10.1016/j.sigpro.2013.10.034>
9. Qobbi, Y., Jarjar, A., Essaid, M., Benazzi, A.: Image encryption algorithm based on genetic operations and chaotic DNA encoding. *Soft comput.* **26**, 5823–5832 (2022). <https://doi.org/10.1007/s00500-021-06567-7>
10. Qobbi, Y., Abid, A., Jarjar, M., Kaddouhi, S. El, Jarjar, A., Benazzi, A.: Adaptation of a genetic operator and a dynamic S-box for chaotic encryption of medical and color images. *Sci. Afr.* **19**, e01551 (2023). <https://doi.org/10.1016/j.sciaf.2023.e01551>

11. Jarjar, M., Hraoui, S., Najah, S., Zenkouar, K.: New technology of color image encryption based on chaos and two improved Vigenère steps. *Multimed. Tools Appl.* **81**, 24665–24689 (2022). <https://doi.org/10.1007/s11042-022-12750-1>
12. Abid, A., Jarjar, M., Benazzi, A., Jarjar, A.: Color image encryption using improved Vigenère method followed by a permutation. In: Ben Ahmed, M., Boudhir, A.A., Santos, D., Dionisio, R., Benaya, N. (eds.) SCA 2022. LNNS, vol. 6, pp. 580–590. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-26852-6\\_54](https://doi.org/10.1007/978-3-031-26852-6_54)



# NBS: An NFT-Based Blockchain Steganography Method

Mustafa Takaoğlu (✉) , Faruk Takaoğlu , and Taner Dursun 

The Scientific and Technological Research Council of Türkiye, Gebze, Kocaeli, Türkiye  
{mustafa.takaoglu, faruk.takaoglu, taner.dursun}@tubitak.gov.tr

**Abstract.** Steganography is the name given to the science of hiding information that has been studied throughout history. With the digitalization of information, traditional physical steganography techniques have been replaced by digital steganography. After the success of the Bitcoin payment system, the potential of blockchain technology, which is the creative technology behind Bitcoin, has been better understood, and it has started to be tried in the field of steganography. In the study, a secured blockchain image-steganography method, which is difficult to detect with steganalysis methods, is proposed. Solana blockchain platform was chosen for the demonstration of the proposed method. In the study, a stego-image was produced from a selected cover-image to hide data by applying the LSB steganography technique, and this stego-image was minted as a Stego-NFT using the Solana platform. The data to be hidden is encrypted with the OTP encryption algorithm. After the transfer of the minted Stego-NFT to the recipient's wallet, the confidential data transmission process is completed with the extraction of the hidden data and then the burning of the transferred NFT. Thanks to the NFT-based image-steganography (NBS) method proposed in the study, the stego-image is only exposed to a stego-only attack during the period between minting and burning. In other words, since only the stego-object is accessible for steganalysis and the stego-NFT is unique, it cannot be detected by steganalysis techniques such as PSNR and histogram analysis. Finally, with these features, the proposed NBS method is one of the first academic studies in this field in the literature.

**Keywords:** Steganography · Blockchain Steganography · LSB · Image Steganography · Blockchain

## 1 Introduction

The word steganography is derived from the Greek words *stegos* and *grafia*, meaning secret writing [1]. As the name suggests, it is the science of secretly sending information to each other that the recipients want to transmit without being noticed by others. The first examples of studies using physical steganography techniques before the digitalization of data were carried out by tattooing the skull of a slave in Aydın/Didim, located in today's Republic of Türkiye, in 500 BC [2]. With the development of computer technology and computer networks, information/data has been digitalized, and confidential data transmission has begun to be carried out with digital steganography methods [3].

Hidden data transmission is carried out by many organizations or individuals, especially intelligence agencies and criminal organizations (such as terrorist organizations, and drug organizations) using steganography techniques for different purposes [4]. The detection of hidden data using steganography is of great importance. Because the importance of hidden data has a lifetime. Most of the time, this data's importance has a very limited lifetime, and if it is not detected or done late, it carries the risk of facing very devastating consequences (such as terrorist attacks). For this reason, various steganalysis tools that can be applied globally are developed and used for the detection of every method of steganography (image, text, video, audio) [5, 6].

In steganography studies, cryptography is generally used. It is aimed that the attacker will not be able to obtain a meaningful result if the hidden data is extracted from a detected stego-image. For this reason, the data is firstly encrypted and then hidden in cover-multimedia by using various steganography techniques. Due to the widespread opinion that it is resistant to the post-quantum era, symmetric encryption techniques (such as Advanced Encryption Standard, AES, and One-Time-Pad, OTP) are generally preferred in studies [7, 8].

The fault-tolerant, decentralized, immutable, and distributed ledger of transaction concept, which is one of the sub-studies of distributed ledger technology and named blockchain technology, was proposed by Satoshi Nakamoto in 2008 [9] and unlike the previous attempts, Digicash [10], BitGold [11], Hashcash [12], and B-Money [13], it has been successful and has gained worldwide acceptance. The Bitcoin protocol proposed by Satoshi has a monolithic architecture. Due to this architecture, it can only meet the need for a decentralized payment method, it causes high energy consumption due to the consensus algorithm it uses, Proof-of-Work, and also causes problems such as long block creation time, high transaction costs, and scalability. And it did not seem possible to use it in different areas except finance [14]. For this reason, the Ethereum protocol, introduced by Vitalik in 2014, has enabled the implementation of blockchain technology in every field with the concept of smart contracts and Ethereum Virtual Machine, unlike Bitcoin [15].

The Ethereum protocol, which was introduced as a world computer but was inspired by Bitcoin, similarly faced energy consumption, speed, and scalability problems, and tried to solve these problems with the open-source Ethereum platform, which is constantly improved and forks (21 forks, which will occur at the end of 2023) thanks to Ethereum Improvement Proposals. The Ethereum protocol, which switched to the Proof-of-Stake consensus algorithm with the Paris (The Merge) Fork in 2022, has become less energy-consuming, and more scalable, unlike Bitcoin, and has come closer to its goal of becoming a world computer [16].

Today, there are other blockchain protocols that offer smart contract development possibilities (Hyperledger Fabric, Solana, Corda etc.). It is possible to generate tokens in blockchain protocols with smart contract capability. Token generation processes in the Ethereum protocol are standardized and are carried out according to the ERC20 standard. Non-Fungible Tokens are produced according to ERC721 and ERC1155 standards. Token generation standards differ on each platform. Regardless of these differences, people's interest in NFTs has increased due to the financial gains that NFT investors have had, and NFTs have become a widely used reality today. Also, NFTs have been offering

solutions to issues that have been for artists for a long time, such as digital ownership and copyright. Contrary to popular belief, NFTs can be not only used for images as well as movies, songs, and etc. too [17].

Blockchain solutions have been seen as a very good alternative medium for data-hiding techniques of steganography. The following different methods have been seen in the Blockchain Steganography studies in the literature:

Sarkar et al. [18] made an image-steganography called Stego-chain in their work, AES symmetric encryption algorithm was used and the file divided into small frames was transmitted to the receiver via blockchain. It is claimed that the file is obtained from these frames that are combined on the receiver side and the hidden data is extracted. However, when the article is examined, it raises deep doubts about realizing the method they claim. The authors' knowledge of blockchain is limited and there is no satisfactory information about how they applied the method they introduced. Despite all these negativities, it is important because it is one of the limited blockchain steganography studies seen in the literature and it is published in an indexed journal. Another similar study was done by Mohsin et al. [19] It has been seen that the researchers who claim that the novelty in their work is in the particle swarm optimization part, shared some inaccurate information in their work. In addition, considering the applicability of the proposed method in terms of blockchain, it does not seem possible to work properly. Similarly, it is important because it is one of the first blockchain steganography studies.

In the study of Basuki and Rosiyadi [20], they proposed a hybrid solution using the Ethereum protocol. They have hidden the encrypted message in the cover-image and proposed a different solution by sending the parameters they keep in the blockchain with the method they introduced as transaction steganography.

With the method named BLOCCE proposed by Partala [21], he proposed blockchain steganography by hiding 1 bit of data in each transaction. The Least Significant Bit technique was used in the study. The only minus of the introduced method is that it requires 8 transactions per byte. However, theoretically, it is one of the most consistent methods among the proposed blockchain steganography based on transactions on the blockchain platform.

Xu et al. [22] proposed to do steganography at the miner level on a public blockchain. Being a miner and gaining the right to write blocks in protocols such as Bitcoin and Ethereum is quite a challenge. In addition, it is not acceptable for a miner who won the block writing award to perform steganography on transactions and miners cannot record data contrary to the block structure introduced in the genesis block. For such reasons, the proposed blockchain steganography is a very theoretical approach and it does not seem possible to implement it.

Giron et al. [23] proposed a steganalysis method for examining public blockchain platforms. In their detailed analysis of Bitcoin and Ethereum blockchains, they did not find any signs of steganography. Innovative studies are needed in this area.

Zichi et al. [24] shared that they suggested a low detectability steganography method suitable for NFTs in their study. To achieve this, enrichment was made in the cover-image (NFT). They shared that the detection rate of this stego-image, after which steganography was applied, was lower.

## 2 Preliminaries

### 2.1 Steganography

Digital steganography is performed using different techniques in the image, text, video, and audio environments. Although these hiding techniques differ according to the type of steganography, there are usually four domains. These are the spatial domain, the transform domain, the spread spectrum domain, and the model-based domain [2].

Cover-image files used in digital image steganography can be in different formats such as PNG, JPEG, BMP, GIF, and TIFF. The same applies to video (MPEG, AVI), audio (MP3, MP4, MPEG, WAV), and text (TXT, PDF, DOC, XLS) steganography techniques. Furthermore, in image steganography, another image or text data can be hidden within a cover image. The type of data to be hidden within the cover multimedia depends entirely on the planned steganography scenario [6].

In this study, image steganography has been performed in the spatial domain using the LSB (Least Significant Bit) technique. In the spatial domain, the data to be hidden is directly embedded in the cover multimedia. The main techniques that stand out in the spatial domain are LSB, Pixel Value Differencing (PVD), and Binary Pattern Complexity (BPC). LSB is a widely preferred technique just because it's fast that is easy to apply [25].

With the LSB technique in Image Steganography, data can be hidden by changing the least significant bit, the last two or three bits of mono-color, or each RGB channel pixel. In the literature, one or two-bit LSB operations are generally preferred. Performing data hiding with more than two bits on an image leads to noticeable distortions [6].

Another important factor to remember is that after the stego-image is produced, distortions that may occur on the stego-image due to various reasons such as compression, scaling, resizing, rotation, etc. can prevent the extraction of the hidden data. To achieve successful digital steganography, imperceptibility (undetectability), security, payload capacity, and robustness should be provided at a certain level [4, 5].

### 2.2 One-Time-Pad (Vernam) Symmetric Encryption Algorithm

Symmetric and asymmetric algorithms are used in encryption. In the post-quantum era, symmetric encryption algorithms are considered secure, and therefore, many steganography studies use symmetric encryption algorithms such as Advanced Encryption Standard (AES), One-Time-Pad (OTP), Blowfish, Triple Data Encryption Algorithm (3DES). In the OTP encryption algorithm, real randomly generated keys are desired to be used. However, true randomness cannot be fully provided by software programs. Therefore, devices that produce keys by imitating various natural events are used. Additionally, in the literature, the MP5 hash algorithm is used to strengthen the random key generation process. In the OTP algorithm, a key consisting of random numbers or characters is generated, which is equal in size to the data to be encrypted. Using this key, the data is processed with XOR logical gate. The encrypted data reaching the recipient is opened using the same key. The generated secret keys must be used only once. A different real random key is generated for each encryption process. Another important issue is key distribution. No matter how complex the encryption algorithm you use is, it is meaningless

if the key is not protected. Various methods can be used for key distribution (such as the Diffie-Hellman key exchange protocol). However, since secure key sharing is outside the scope of this study, it is not described in detail [7].

### 2.3 Blockchain Technology

Blockchain technology is an innovative solution that emerged with the success of the Bitcoin protocol, which was shared in 2008 as a whitepaper titled “Bitcoin: A Peer-to-Peer Electronic Cash System”. The Bitcoin protocol is seen to have a highly secure structure. Contrary to popular belief, no encryption algorithm is used in the Bitcoin protocol. The system, which is created using hash and digital signature algorithms, enables transactions to become immutable when they are written to blocks. Concepts such as network protocols, Merkle trees, SHA256 hash algorithm, secp256k1 elliptic curve, Proof-of-Work (PoW) consensus algorithm, block structure (timestamp, nonce, Merkle root, difficulty index, version, previous block hash), UTXO, mining, and mining reward form the basis of the Bitcoin protocol [25].

In Bitcoin, a transaction process takes place as follows: The sender determines the amount of BTC they want to send from their wallet to the receiver’s public address and submits a sending request. This transaction request is then sent to the memory pool, which is a queue of transaction requests operated by miners. The transaction requests are then sent to the network for confirmation to be added to the blockchain in block form by the miner who has won the mining reward. If the majority of nodes in the network accept the transaction, the block containing the transaction information is added to the blockchain, and the transaction is completed. The blocks are linked to each other using the SHA256 hash algorithm, and all blocks except for the first block called the genesis block, contain the hash value of the previous block. Thus, any attempt to make changes to a block will be detected and rejected since any change will result in a different hash value due to the avalanche effect of the SHA256 based linkage between blocks. This data architecture, which is obtained by linking the blocks to each other using the SHA256 hash algorithm, is called a ledger. Similar to pages in a ledger, a new block is added when a block is filled, and these blocks are continuously created and added to the chain. Therefore, the Bitcoin protocol is referred to as an immutable (tamper-proof) ledger of transactions. Additionally, the Bitcoin protocol is a fault-tolerant system that solves the Byzantine general’s problem, which was introduced by Leslie Lamport [26].

As outlined in the Bitcoin whitepaper, the protocol only provides a peer-to-peer payment system solution. However, its implementation presents much larger potential. Therefore, the Ethereum protocol was introduced by Vitalik to enable the immutable, distributed, decentralized, fault-tolerant system offered by Bitcoin to be used in non-financial areas by becoming a world-state computer. With the Ethereum protocol, users can achieve the solutions they need in non-financial areas through the execution of programs coded for specific tasks, called smart contracts, on the Ethereum Virtual Machine. The technology offered by Bitcoin as a static solution with its monolithic architecture has become applicable in many areas with the Ethereum protocol. This technology is now commonly referred to as blockchain [25].

### **Non-Fungible Tokens (NFTs)**

A Non-Fungible Token (NFT) is utilized for the unique identification of something or someone. NFTs use distinctive identification codes and metadata to differentiate them from each other on a blockchain platform, such as Ethereum, Solana, and others. NFTs can be utilized on platforms that offer collectible items, access keys, lottery tickets, concert tickets, sports matches, blockchain games, and so on. There are several differences between fungible and non-fungible tokens. NFTs are unique, indivisible, and non-convertible, which distinguishes them from fungible tokens. NFTs can be transferred between wallets but cannot be converted into a different digital asset.

When starting an NFT project, the first step is to determine what type of NFT will be produced (such as an image, song, GIF, etc.). Then, the appropriate blockchain protocol is chosen, and a wallet is created that is compatible with the chosen protocol. A smart contract is coded according to the NFT standard of the chosen platform and uploaded to the blockchain protocol. The required transaction fee is paid, and the desired NFT is minted. The produced NFT can be kept in the user's wallet or listed for sale on an NFT marketplace. Additionally, the produced NFT can be transferred to another compatible wallet. If desired, NFTs can be burned, which is the mechanism for removing a token or NFT from circulation. The burning mechanism is irreversible, and it requires payment of a fee for the transaction to the burn address of the platform being used [27].

NFTs can be entirely on-chain, which means that both the image and its metadata are stored on the blockchain, or off-chain, which means that some of the NFT is stored off the blockchain. In many NFT projects, images are stored off-chain on users' servers or in services such as the InterPlanetary File System (IPFS) or BigchainDB, etc. Various methods are used in the production of NFTs. For example, in NFTs called Fractional NFTs, a picture/song/land deed is divided into pieces and each piece is produced and sold as an NFT. Another method is Generative Art NFTs, which are generated by computer programs. NFTs can also be produced on the Bitcoin blockchain with the Ordinals protocol. The Ordinals protocol is a solution that allows tracking of Satoshi's, the smallest value of Bitcoin, by assigning them a serial number, called inscription [27].

### **Solana Protocol**

Solana blockchain network, launched in 2020, is an innovative layer-1 blockchain protocol developed with the scalability limitations of established blockchains like Ethereum in mind. It is an open-source and public blockchain protocol. Due to its high scalability, low transaction times, and costs, the Solana protocol is widely preferred for NFT solutions. The Solana platform has smart contract capabilities, referred to as programs. Programs can be developed in any language. Solana has SeaLevel parallel smart contract run-time. In addition to Proof-of-History, the Tower BFT algorithm, Turbine block propagation protocol, and Gulf Stream memory pool-less transaction forwarding protocol enable the Solana protocol to achieve high speeds, and its horizontally-scaled accounts database makes it more scalable than other alternatives. Solana Program Library (SPL) is used to generate tokens in the Solana protocol. SPL creates a standardized interface for the minting, issuing, transferring, and destruction of Solana-compatible tokens, similar to the ERC standards in Ethereum [28].



### 3 Proposed Method

With the proposed NBS method, blockchain technology, OTP symmetric encryption algorithm, generative art NFTs, and traditional image-steganography technique (1-bit LSB) are used in the most optimal way, suggesting a hard-to-detect and highly secure blockchain steganography method that hides text data. The general structure of the proposed NBS method is shared in Fig. 1.

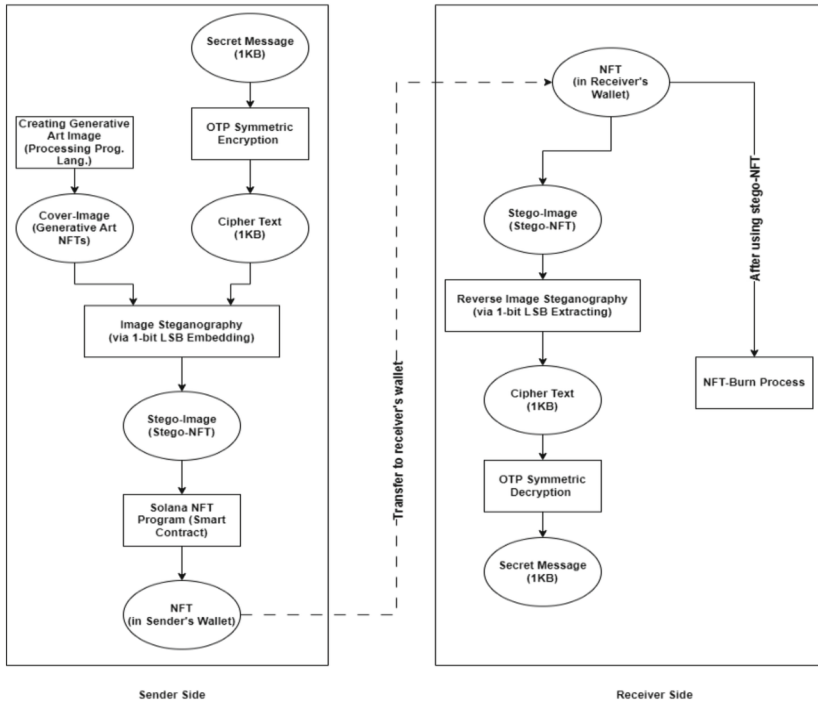
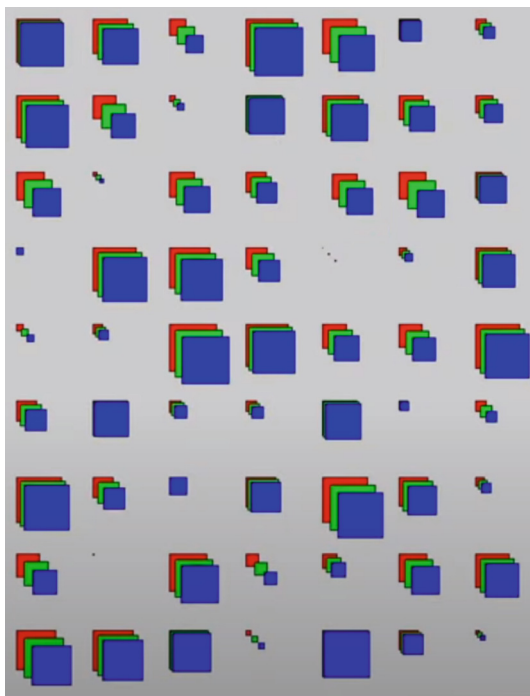


Fig. 1. Structure of NBS method.

As can be seen in Fig. 1, cover-images used in the proposed NBS method were produced with the Generative Art NFT code developed using the Processing programming language. A 1KB.txt file was created as the secret message. A random 1KB secret key was produced to be used in the OTP encryption algorithm. By using the secret message and the secret key with the OTP algorithm (XOR), the ciphertext is produced. The ciphertext is hidden in the generated cover-image with the 1-bit LSB technique. The resulting stego-NFT is shared in Fig. 2.

Two Solana wallets were created for testing purposes and money (dummy Sol) was loaded into these wallets. The prepared NFT mint program (smart contract) has been uploaded on Solana's Devnet. The NFT was created by sending the generated stego-NFT's link (IPFS URI) to the Solana smart contract by sender. Then the created NFT has been transferred to the sender's wallet then the sender sent its NFT to the receiver's



**Fig. 2.** Stego-NFT.

wallet (transaction fee paid by the sender with Sol). On the Receiver-side, the stego-image (stego-NFT) was taken from the transferred NFT and subjected to reverse LSB processing. By using the OTP secret key which is shared between the parties before, the ciphertext was deciphered and the secret message was accessed. Simultaneously, the NFT (stego-NFT) received in the wallet was destroyed using the SPL Token Burn function (We did not pin NFTs on IPFS therefore Stego-NFT will be erased on IPFS eventually). All the codes of the proposed NBS method have been uploaded to the <https://github.com/mustafatakaoglu> page and the code will be shared with those who want to examine it.

## 4 Conclusion

With the NFT-based Blockchain Steganography (NBS) method, blockchain image steganography is carried out using NFTs. Encrypting the secret message with the One-Time-Pad (OTP) symmetric encryption algorithm, which is resistant to the post-quantum era, ensures that the hidden data cannot be accessed in case of the detection of stego-NFT. Stego-NFT is obtained as a result of the encrypted secret message hidden in the cover-image with the Least Significant Bit (LSB) technique. This stego-NFT is then minted as NFT on the Solana platform. The generated NFT is transferred to the receiver's wallet and the secret message is obtained by following the steps suggested in the NBS method,

and then the transferred NFT is burned by the receiver. Since burned NFTs are not accessible, steganalysis will no longer be possible on them. Since generative art NFTs are produced in the proposed NBS method, it is not possible to detect the data hidden in these unique images by any steganalysis method. Also, burning NFTs keeps users even safer. Since the implementation of the NBS method with the same wallets and smart contracts all the time may attract attention over time, it is beneficial to change the wallets at certain intervals (or one-time-wallet for each transaction) and to renew smart contracts similarly. The Solana protocol was preferred for the NBS method because the transactions are fast and cheap. However, the NBS method is platform agnostic. It can be used in all protocols with smart contract capability. Apart from Generative Art programs, traditional images can also be used as cover-images. In this case, different steganography techniques or improvement techniques such as Optimal Pixel Adjustment Process (OPAP) can be used to make it difficult to detect with steganalysis methods.


## References

1. Sönmez, F., Takaoğlu, F., Kaynar, O.: Ideal steganography scenario: calculation of capacities of carrier images, OPA method in frequency based steganography. *ACTA INFOLOGICA* **2**(1), 12–21 (2018)
2. Şahin, F., Çevik, T., Takaoğlu, M.: Review of the literature on the steganography concept. *Int. J. Comput. Appl.* **183**(2), 38–46 (2021)
3. Su, W., Ni, J., Hu, X., Fridrich, J.: Image steganography with symmetric embedding using Gaussian Markov random field model. *IEEE Trans. Circuits Syst. Video Technol.* **31**(3), 1001–1015 (2021)
4. Takaoğlu, F., Takaoğlu, M.: Printer steganography, yellow dot analysis - a mini survey. *ArtGRID-J. Archit. Eng. Fine Arts* **1**(1), 25–35 (2019)
5. Takaoğlu, F., Takaoğlu, M.: Today's validity of printer steganography and yellow dot analysis. *e-J. New Media* **4**(3), 176–184 (2020)
6. Takaoğlu, F., Takaoğlu, M.: Hiding image and text data with DCT and DWT techniques. *J. Istanbul Aydın Univ.* **12**(3), 189–200 (2020)
7. Takaoğlu, M., Özyavaş, A., Ajlouni, N., Takaoğlu, F.: Highly secured hybrid image steganography with an improved key generation and exchange for one-time-pad encryption method. *Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi* **23**(1), 101–114 (2023)
8. Hassaballah, M., Hameed, M.A., Awad, A.I., Muhammad, K.: A novel image steganography method for industrial internet of things security. *IEEE Trans. Industr. Inf.* **17**(11), 7743–7751 (2021)
9. Bitcoin: a peer to peer electronic cash system. <https://bitcoin.org/bitcoin.pdf>. Accessed 26 Mar 2023
10. Chaum, D.: Blind signatures for untraceable payments. *Adv. Cryptol. Proc. Crypto* **82**(3), 199–203 (1983)
11. Bit gold. <https://web.archive.org/web/20151121081112/http://unenumerated.blogspot.com/2005/12/bit-gold.htm>. Accessed 26 Mar 2023
12. Hash cash. <http://www.hashcash.org/papers/announce.txt>. Accessed 26 Mar 2023
13. B-Money. <https://nakamotoinstitute.org/b-money/>. Accessed 26 Mar 2023
14. Takaoğlu, M., Özer, Ç., Parlak, E.: Blockchain technology and possible implementation areas in Turkey. *Uluslararası Doğu Anadolu Fen Mühendislik ve Tasarım Dergisi* **1**(2), 260–295 (2019)
15. Ethereum whitepaper. <https://ethereum.org/en/whitepaper/>. Accessed 26 Mar 2023

16. Ethereum history. <https://ethereum.org/en/history/>. Accessed 26 Mar 2023
17. Zubaydi, H.D., Varga, P., Molnár, S.: Leveraging blockchain technology for ensuring security and privacy aspects in internet of things: a systematic literature review. *Sensors* **23**, 788 (2023)
18. Sarkar, P., Ghosal, S.K., Sarkar, M.: Stego-chain: a framework to mine encoded stego-block in a decentralized network. *J. K. S. U. Comput. Inf. Sci.* **34**(8), 5349–5365 (2020)
19. Mohsin, A.H., et al.: PSO–blockchain-based image steganography: towards a new method to secure updating and sharing COVID-19 data in decentralised hospitals intelligence architecture. *Multimed. Tools Appl.* **80**, 14137–14161 (2021)
20. Basuki, A.I., Rosiyadi, D.: Joint transaction-image steganography for high capacity covert communication. In: *Proceedings of the 2019 International Conference on Computer, Control, Informatics and Its Applications, IC3INA 2019*, pp. 41–16. IEEE, Tangerang (2019)
21. Partala, J.: Provably secure covert communication on blockchain. *Cryptography* **2**(3), 18 (2018)
22. Xu, M., Wu, H., Feng, G., Zhang, X., Ding, F.: Broadcasting steganography in the blockchain. In: Wang, H., Zhao, X., Shi, Y., Kim, H.J., Piva, A. (eds.) *IWDW 2019*. LNCS, vol. 12022, pp. 256–267. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-43575-2\\_22](https://doi.org/10.1007/978-3-030-43575-2_22)
23. Giron, A.A., Martina, J.E., Custódio, R.: Steganographic analysis of blockchains. *Sensors* **21**(12), 4078 (2021)
24. Zichi, W., Guorui, F., Xinpeng, Z.: Steganography in NFT images. *Chin. J. Netw. Inf. Secur.* **8**(3), 18–28 (2022)
25. Takaoglu, M., Özyavaş, A., Ajlouni, N., Alshahrani, A., Alkasasbeh, B.A.: Novel and robust hybrid blockchain and steganography scheme. *Appl. Sci.* **11**(22), 10698 (2021)
26. Lamport, L., Merz, S.: Specifying and verifying fault-tolerant systems. In: Langmaack, H., de Roever, W.P., Vytupil, J. (eds.) *FTRTFT ProCoS 1994*. LNCS, vol. 863, pp. 41–76. Springer, Heidelberg (1994). [https://doi.org/10.1007/3-540-58468-4\\_159](https://doi.org/10.1007/3-540-58468-4_159)
27. Radermecker, A.-S.V., Ginsburgh, V.: Questioning the NFT “revolution” within the art ecosystem. *Arts* **12**(1), 25 (2023)
28. Solana: A new architecture for a high performance blockchain. <https://solana.com/solana-whitepaper.pdf>. Accessed 29 Mar 2023



# MLP Neural Network Based on PCA and $K$ -means Clustering for PM<sub>2.5</sub> Forecasting

Diego Velez, Santiago Santa, and Gustavo Patino<sup>(✉)</sup> 

SISTEMIC Group, Engineering School, University of Antioquia, Medellín, Colombia  
{diego.velez1,santiago.santa,adolfo.patino}@udea.edu.co  
<https://shorturl.at/fjsAB>

**Abstract.** Air pollution poses a significant environmental challenge, adversely affecting the health of millions worldwide. Consequently, accurate prediction of pollutant levels has become increasingly crucial to prevent and mitigate the negative impacts of air pollution. This research introduces a Python-based artificial neural network algorithm for predicting PM<sub>2.5</sub> levels in Medellín, Colombia, leveraging meteorological and emission data. The model utilizes a Multilayer Perceptron Neural Network, incorporating principal component analysis (PCA) and  $K$ -means clustering to determine the optimal number of hidden layers and neurons. Additionally, trend and correlation analyses were conducted to identify the most relevant predictors by examining the relationship between available variables and the target variable (PM<sub>2.5</sub>). Model performance is assessed using Mean Square Error and Mean Absolute Error.

**Keywords:**  $K$ -means Clustering · Leaky ReLU function · Multilayer Perceptron (MLP) · PM<sub>2.5</sub> · Principal Component Analysis (PCA)

## 1 Introduction

Particulate Matter 2.5, also known as PM<sub>2.5</sub>, is a category of air pollutants with an aerodynamic diameter of less than 2.5 micrometers ( $\mu\text{m}$ ) [1]. These pollutants can have a detrimental effect on human health when individuals are exposed to them constantly or for prolonged periods, as they can enter the lungs and cause respiratory and/or cardiovascular conditions [2]. Numerous research projects have been conducted worldwide to analyze the causes of high levels of particulate matter and their effects on populations. For instance, in the city of Hong Kong, X. Li et al. [3] analyzed the impact of certain meteorological variables on the concentration of PM<sub>2.5</sub> by performing a correlation analysis using a dataset with information from January to December 2013. The results showed that only the meteorological variable pressure had a similar trend to that of PM<sub>2.5</sub>. In contrast, the variables of temperature, rainfall level, relative humidity, wind speed, and direction presented an inverse relationship with respect to particulate matter.

A similar study was carried out in the city of Bogotá by Franceschi, F et al. [4]. The authors used Principal Component Analysis (PCA) [5] to determine the most influential meteorological variables on air pollution in Bogotá for a dataset between 2010 and 2015. They measured the data at 13 local monitoring stations. Additionally,  $K$ -means

---

Supported by CODI, University of Antioquia, Colombia.

clustering analysis [6] was used to group the PCA results and configure them as input data for two multilayer perceptron (MLP) neural networks [7] that predicted the level of PM10 and PM2.5 in the city. The study found that using the clustering results as input variables improved the performance of the PM10 and PM2.5 level prediction models. In Tianjin, a city located in northern China, Gao et al. [8] employed a combination of principal component analysis (PCA) [5] and an artificial neural network (ANN) to predict PM2.5 levels, as one of four models. The authors found that the use of PCA and ANN in combination produced the most accurate results, with prediction errors of R2 [9, 10] at 0.99 and RMSE [10] less than 15.

Two other related works were conducted by authors Choi, S et al. and Liu, W et al. in which prediction of PM2.5 levels in the cities of Seoul (South Korea) and Beijing (China), respectively, was performed. Choi, S et al. [11] found that the application of PCA in deep learning time series prediction can contribute to improved practical performance even with a small number of observations.

On the other hand, Rachmatullah, M et al. [12] conducted a study to determine the number of deep layers and neurons in each layer of a MLP Neural Network for predicting wind speed based on relevant variables. The authors utilized Principal Component Analysis (PCA) [5] and *K*-means clustering [6] to design the ANN. The number of deep layers in the network was determined by the number of components contributing the most variance obtained through PCA analysis, while the optimal number of neurons for each layer was determined by the number of data groups obtained through *K*-means clustering analysis.

This article presents a computational algorithm to predict the PM2.5 index in certain areas of Medellin (Colombia) with several days or weeks in advance. The algorithm performs a correlation analysis between data collected from the SIATA website [13], which includes measurements of the PM2.5 index and other pollutant and meteorological variables. This analysis helps to identify the variables with the highest correlation with the PM2.5 index, which are used as input variables in a MLP neural network [7] to predict the PM2.5 index. To determine the optimal architecture of the ANN, PCA analysis [5] and *K*-means clustering [6] were used to define the number of deep layers and the amount of neurons in each layer, respectively. Consequently, the present study aims to achieve the following contributions in developing neural network models for PM2.5 prediction:

1. Explore methodologies for configuring a Multi-layer Perceptron Neural Network tailored for PM2.5 prediction.
2. Introduce an efficient prediction accurate for algorithm of PM2.5 behavior, enhancing understanding and enabling proactive pollution control measures.
3. Contribute to science and technology, particularly within the local community, by utilizing advanced neural network models to improve air quality standards and promote a healthier living environment.

The paper is structured as follows: Sect. 2 explains the Methodology and Analysis carried out in this project with respect to data collection, correlation analysis and trend analysis. There, two different Neural Network models based on PCA and *K*-means Clustering are also described, considering the results of several simulations and their corresponding performance comparison. Finally, in Sect. 3 the conclusions and future work are discussed.

## 2 Methodology and Analysis

The block diagram presented in Fig. 1 describes the algorithm proposed in this article, the most relevant steps or stages of which are described below:

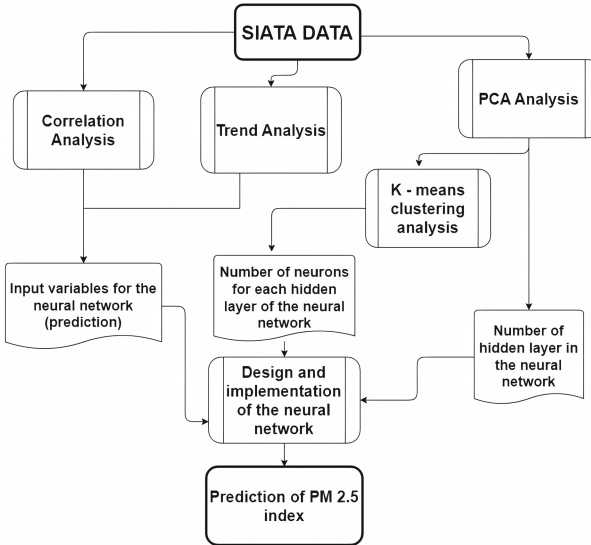


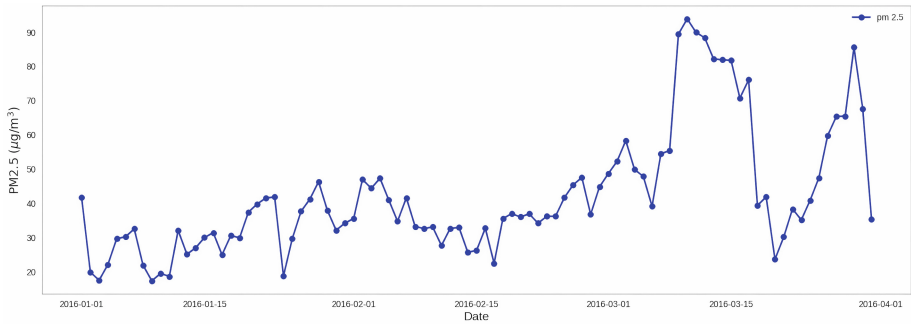
Fig. 1. Algorithm for the prediction of the PM2.5 level.

### 2.1 Data Collection or Acquisition

The “*Sistema de Alerta Temprana de Medellín y el Valle de Aburrá*” (SIATA) [13] has a network of sensors distributed throughout Medellín city, which provide measurements of Criteria air pollutants<sup>1</sup> and meteorological variables 24 hours a day. The data can be downloaded through its website, so that such data represents a time series for each of the measured variables, where the rows represent time and the columns represent data.

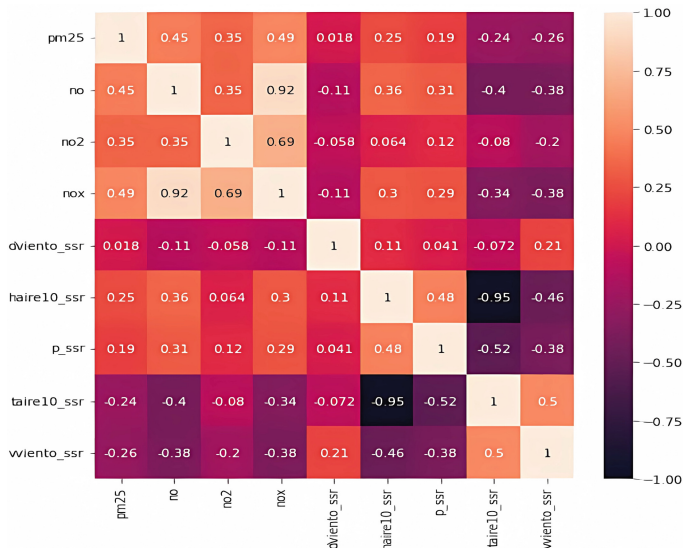
The sensor network of SIATA [13] is distributed throughout different locations in the city, and through the analysis of results obtained using a data cleaning program, station number 28, called “*Casa de Justicia de Itagiüt*” by SIATA, was defined as the station to be used for the PM2.5 prediction process, as it provides the highest amount of variables with valid data, from January 1, 2016 to December 31, 2019. The process of visualization, data analysis, and other aspects related to the development of prediction algorithms were performed in Python. Figure 2 shows the PM2.5 index (daily average) for the first three months of 2016, obtained using the Pandas and Matplotlib libraries in Python [15]. The following sections describe the data analysis processes performed on the available predictor variables.

<sup>1</sup> Criteria air pollutants are substances for which acceptable levels of exposure can be determined and for which an ambient air quality standard has been set [14].



**Fig. 2.** PM2.5 particulate matter index.

**Correlation Analysis:** To identify predictor variables with a linear relationship to the target variable (PM2.5), a correlation analysis was conducted by obtaining the Pearson coefficient [16] for each variable and representing them in a matrix. The correlation matrix displays the correlation value for each predictor variable through a numerical value and color. Variables with a correlation value close to 1 (positive relationship) or  $-1$  (negative relationship) are represented with a light or dark color, respectively. The Seaborn library in Python [15] was used to create this matrix. Figure 3 displays the correlation matrix for the analyzed dataset.



**Fig. 3.** Correlation matrix for the SIATA dataset - Year 2016.



Based on the analysis of Fig. 3, it can be concluded that  $NO_x$ ,  $NO$ ,  $NO_2$ , wind direction, air humidity, and pressure have a positive relationship with the PM2.5 index, while air temperature and wind speed have a negative relationship with PM2.5.

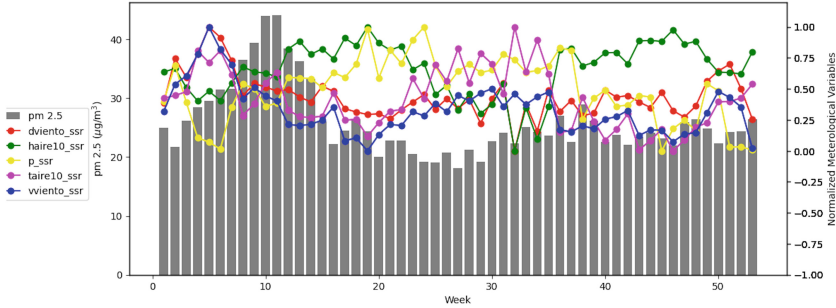


Fig. 4. Weekly trend of PM2.5 concentration and meteorological variables.

**Trend Analysis - Meteorological Variables:** Trend analysis was performed to investigate the relationship between PM2.5 and meteorological variables at monthly and weekly intervals using SIATA data [13]. Figure 4 shows the graph for the weekly trend analysis. In addition, the results for the monthly trend analysis between meteorological variables and the PM2.5 index are shown in Fig. 5.

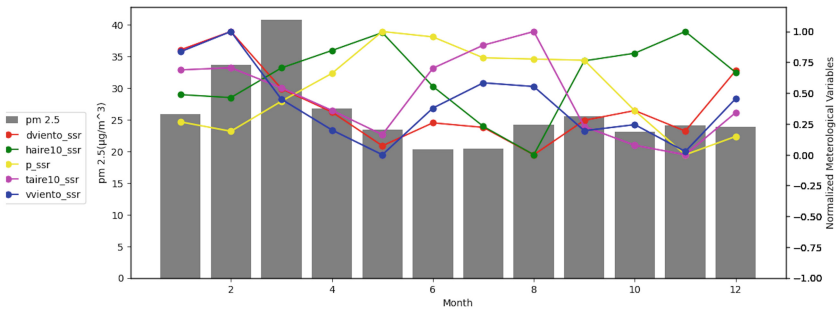


Fig. 5. Monthly trend of PM2.5 concentration and meteorological variables.

After analyzing Fig. 4 and Fig. 5, it can be concluded that the wind direction variable has a similar trend to the PM2.5 index, whereas the pressure, wind speed, and air temperature variables exhibit an inverse trend. Thus, the results of the correlation and trend analysis indicate that the optimal variables for predicting the PM2.5 index using an ANN are  $NO_x$ ,  $NO$ ,  $NO_2$ , wind speed, air temperature, and pressure. These pollutant and/or meteorological variables will serve as the inputs for the neural network, determining the number of neurons required in the input layer for predicting the PM2.5 index.

## 2.2 Designing the Neural Network Using PCA and $K$ -means Clustering

The dataset utilized in this study comprises various time series, which are sequences of data points ordered in time, where time is typically the independent variable, and the objective is to forecast the value of a specific variable of interest in the future. Mathematical models such as ARIMA [17] have been effective in short-term prediction of high-frequency time series. However, ANN are considered more resilient, particularly when capturing complex relationships that exhibit non-linear behaviors [18]. To predict the PM2.5 index, a MLP neural network was employed. The design of this type of ANN involves determining the number of deep layers, and the number of neurons associated with each layer. The number of neurons in the input layer is determined by the number of predictor variables used for the prediction, while the number of neurons in the output layer is defined by the number of variables to be predicted. Various methods can be used to determine the number of deep layers and the number of neurons in each layer. For instance, in the Sartoti method [19], it is suggested to use a single deep layer, such that the number of neurons is determined by the relation:

$$\text{Number of neurons} = N - 1 \quad (1)$$

where  $N$  corresponds to the number of inputs of the ANN, i.e., the number of predictor variables used for the prediction. On the other hand, the Tamura and Tateishi method [20] suggests using two deep layers, whose number of associated neurons corresponds to the relation:

$$\text{Number of neurons} = N/2 + 3 \quad (2)$$

To determine the optimal ANN used in this article, PCA [5] and  $K$ -means clustering [6] analyses were carried out.

**Principal Component Analysis - PCA:** It is a statistical technique used to simplify the complexity of sample spaces with high-dimensional data while retaining their most relevant information. It reduces the number of dimensions of a sample space (variables) from  $p$  to a smaller number of principal components  $z$ , where  $z < p$ , that preserve the same information as the original variables [5]. These principal components can then be used to describe or analyze the phenomenon represented by the variables using fewer data points or information. There are various techniques to determine the optimal number of components for a dataset, such as the explained variance percentage [5], the Kaiser criteria [21], and the Jolliffe modification [22]. The explained variance percentage method is the most straightforward, which involves setting a threshold percentage, such as 90%, and selecting principal components in a successive manner until the threshold is exceeded [23].

*Determining the Number of Deep Layers of Our Neural Network:* Using the explained variance percentage method [5], Fig. 6 indicates that for this case study, the appropriate number of components is 4, since the first four principal components explain 88% of the variance.

**$K$ -means Clustering Analysis:** The term “*clustering*” refers to a wide range of unsupervised techniques whose purpose is to find groups (“*clusters*”) within a set of observations [6]. The “*K-means clustering*” method involves grouping observations into  $K$

different clusters, where the number  $K$  is predetermined by the analyst before executing the algorithm. The goal of “ $K$ -means clustering” is to identify the optimal  $K$  clusters, where the best cluster is defined as the one with the smallest internal variance (i.e., intra-cluster variation) [6].

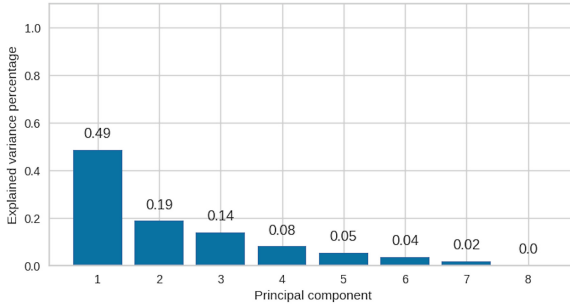


Fig. 6. Explained variance percentage for each component - PCA Analysis.

*Definition of the Number of Neurons for Each Hidden Layer:* This number was determined using the  $K$ -means clustering analysis, such that said number depends on the total number of clusters found by the algorithm. Figure 7 displays the results of the  $K$ -means clustering for the first (PC1) component obtained through PCA analysis. The figure utilizes the *elbow* method [24] to determine the appropriate number of clusters  $K$ . The *elbow* method involves running the  $K$ -means clustering on the data set for a range of  $K$  values and calculating the sum of squared errors (SSE)<sup>2</sup> for each value  $K$ .

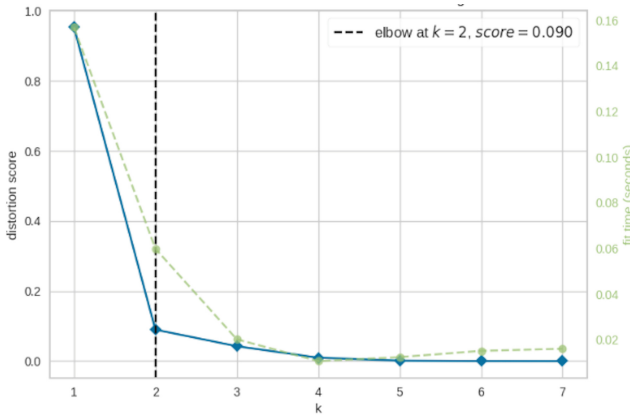


Fig. 7. Distortion Score Elbow -  $K$ -means clustering analysis.

In Fig. 7, the graph shows the behavior of the SSE error value (y-axis) as the value of  $K$  (x-axis) increases, represented by a continuous blue line. The shape of the line can be

<sup>2</sup> It is the sum of squared differences between each observation and the mean of its group [25].

visualized as an “arm”, where the “elbow” indicates the appropriate value of  $K$  for the  $K$ -means clustering. The green dashed line shows the amount of time needed to train the clustering model for each value  $K$ . Finally, the black dashed line represents the *elbow* that corresponds to the optimal value of  $K$ , which was found using the *knee point detection algorithm* [26]. Therefore, the appropriate number of neurons for the first hidden layer (represented by the PC1 component) is **2**. Using a similar procedure as described above, the number of neurons was determined for the three remaining hidden layers, represented by the PC2, PC3, and PC4 principal components. The ANN configuration takes into account that the principal components with the highest variance correspond to the hidden layers closer to the output [12]. Table 1 shows the final configuration for the hidden layers of our neural network.

**Table 1.** Configuration for the deep layers of the neural network.

Principal Component	Number of Deep Layer	Number of Neurons
PC4	1	3
PC3	2	3
PC2	3	2
PC1	4	2

**Selection of Activation Functions and More Neural Network Parameters:** Activation functions are utilized in ANN to convert the input signal from a neuron in a given layer, to an output signal that functions as input to the following layer. These functions enable the neuron to decide whether or not to activate by computing the weighted sum and adding a bias. This introduces nonlinearity into the output of a neuron [27], resulting in outputs ranging from  $[0,1]$  or  $[-1,1]$ . Some of the most commonly used activation functions in deep layers include sigmoid, hyperbolic tangent, and ReLU (short for Rectified Linear Unit) [28]. In this article, the activation function selected was the Leaky ReLU function [27], which is an improvement over the ReLU activation function. The linear function was chosen as the activation function for the output layer of the ANN, which is suitable for developing regression models for prediction [29,30]. Table 2 show the parameters defined for the ANN.

**Table 2.** Neural Network Parameters

AF- Deep Layers	AF - Output Layer	Learning rate	Epochs
LeakyReLU	linear	0.001	1000

### 2.3 Training, Prediction, and Validation of the Implemented Neural Network

To validate the prediction model developed in this study, two ANN models were employed to forecast the level of PM2.5. The first model, called **Model 1**, includes

all the available predictor variables from the dataset obtained from SIATA [13] as the input layer. Figure 8 displays the prediction results obtained using **Model 1**. On the other hand, the second model, **Model 2**, only includes the most relevant predictor variables obtained through correlation and trend analysis. The corresponding results are displayed in Fig. 9.

**Comparison of Predictions:** The implemented ANN models use the Mean Squared Error (MSE) [31] as the loss function, and the Mean Absolute Error (MAE) [31] as the metric function. Equations 3 and 4 illustrate the expressions for MSE and MAE, respectively.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{3}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{4}$$

where:

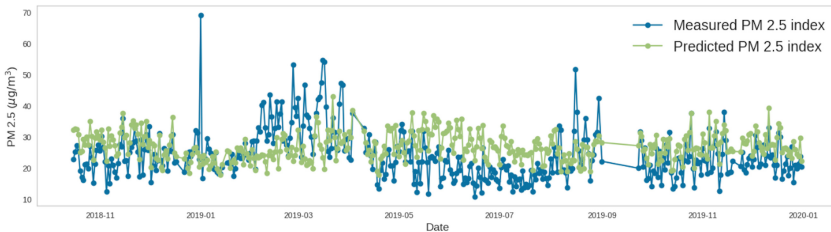
- $n$  is the total number of samples.
- $y_i$  denotes the actual values of the samples.
- $\hat{y}_i$  represents the predicted values for the samples.

To evaluate the performance of these models, Table 3 displays the corresponding values of these functions for each analyzed model.

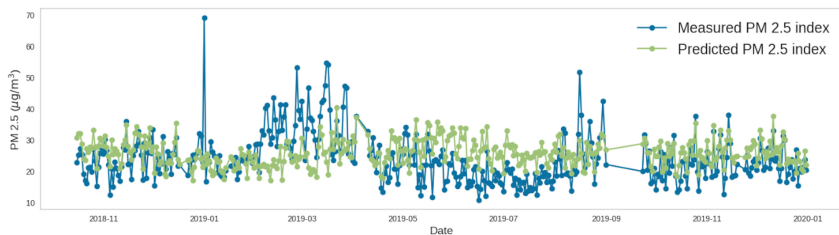
**Table 3.** Comparison of prediction error - implemented neural network models.

Neural network model	MSE	MAE
Model 1	70.6	6.73
Model 2	66.0	6.25

Based on the results in Table 3, it can be inferred that the most accurate prediction model is **Model 2**, which uses as input only the selected predictor variables based on the correlation and trend analysis. Consequently, using as input the variables that are more closely related to the predicted variable in our ANN model improves its performance, reducing the error of the prediction value.



**Fig. 8.** PM2.5 index prediction - **Model 1**.



**Fig. 9.** PM<sub>2.5</sub> index prediction - **Model 2.**

### 3 Conclusions

Trend and correlation analysis of variables are appropriate methodologies for selecting ideal variables for regression model prediction, such that the Leaky ReLU function is suitable for implementing ANN for regression models. Additionally, the implementation of PCA and *K*-means Clustering as methodologies for designing multilayer perceptron ANN establishes an effective and efficient algorithm that enables obtaining low prediction error values, so that the results obtained confirm the validity of the prediction paradigm found through the state-of-the-art analysis used for this research, in which determining the variables that have the greatest relationship with the target variable increases the prediction performance. In future work, including additional variables like PM<sub>10</sub> is expected to improve prediction accuracy. A larger and up-to-date dataset will enhance prediction performance by considering current meteorological conditions. These enhancements will ensure comprehensive and relevant data for more accurate predictions.

### References

1. Culp, J.: Stay Safe in the PM<sub>2.5</sub> (2019). [https://www.chula.ac.th/wpcontent/uploads/2020/01/e-Book\\_Stay-Safe-in-the-PM2.5-EN.pdf](https://www.chula.ac.th/wpcontent/uploads/2020/01/e-Book_Stay-Safe-in-the-PM2.5-EN.pdf). Accessed 1 Dec 2022
2. World Health Org. Ambient Air Pollut.: A Global Assess. of Exposure and Burden of Disease (2016). <https://apps.who.int/iris/handle/10665/250141>. Accessed June 2020
3. Li, X., Feng, Y., Liang, H.: The impact of meteorolog. factors on PM<sub>2.5</sub> variations in Hong K. IOP Conf. Ser. Earth Environ. Sci. **78** (2017)
4. Franceschi, F., Cobo, M., Figueredo, M.: Discovering relationships and forecast. PM<sub>10</sub> and PM<sub>2.5</sub> concentrations in Bogotá, using ANN, PCA and k-means clustering. Atmosph. Pollut. Res. **9**(4), 697–704 (2018)
5. Rodrigo, J.A.: Principal Component Analysis (PCA) and t-SNE. Cienciadedatos.net (2017). [https://www.cienciadedatos.net/documentos/35\\_principal\\_component\\_analysis](https://www.cienciadedatos.net/documentos/35_principal_component_analysis). Accessed 21 Sept 2022
6. Rodrigo, J.A.: Clustering and Heatmaps: Unsuperv. Learn. Cienciadedatos.net (2017). [https://www.cienciadedatos.net/documentos/37\\_clustering\\_y\\_heatmaps](https://www.cienciadedatos.net/documentos/37_clustering_y_heatmaps). Accessed 21 Sept 2022
7. Rodrigo, J.A.: ANN with Python (2021). <https://www.cienciadedatos.net/documentos/py35-redes-neuronales-python.html>. Accessed 11 Oct 2022

8. Gao, S., Zhao, H., Bai, Z., et al.: Combined use of PCA and ANN to improve personal exposure estimates of PM<sub>2.5</sub>: a case study in elderly. *Sci. Total Environ.* **726** (2020)
9. MathWorks. Coefficient of Determination (R-Squared) R2023. <https://www.mathworks.com/help/stats/coefficient-of-determination-r-squared.html>. Accessed 23 Oct 2022
10. Chugh, A.: MAE, MSE, RMSE, Coefficient of Determination, Adjusted RSquared - Which Metric is Better? (2020). <https://medium.com/analyticsvidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-whichmetric-is-better-cd0326a5697e>. Accessed 23 Oct 2022
11. Choi, S., Kim, B.: Applying PCA to deep learning prediction models for predicting PM<sub>2.5</sub>. *Sustainability* **13**(7) (2021)
12. Rachmatullah, M.I.C., Santos, J., Surendro, K.: Determining the number of layers and hidden neurons in a NN. for wind speed prediction. *PeerJ Comput. Sci.* **7**(e724) (2021)
13. SIATA. SIATA - Sistema de Alerta Temprana del valle de Aburrá. [https://siata.gov.co/siata\\_nuevo/](https://siata.gov.co/siata_nuevo/). Accessed 29 Aug 2022
14. Center for Disease Control and Prevention - CDC. Air Pollutants (2022). <https://www.cdc.gov/air/pollutants.htm>. Accessed 01 May 2023
15. Python. Welcome to Python.org. <https://www.python.org/>. Accessed 29 Aug 2022
16. Ortega, C.: What is Pearson's correlation coefficient? (2019). <https://www.questionpro.com/blog/es/coeficiente-de-correlacion-de-pearson/>. Accessed 31 Oct 2022
17. Chávez Quisbert, N.: ARIMA Models (1997). [http://www.scielo.org.bo/scielo.php?script=sci\\_arttext&pid=S2077-33231997000100005&lng=es&tlng=es](http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S2077-33231997000100005&lng=es&tlng=es). Accessed 12 Oct 2022
18. Mercado Polo, D., Pedraza Caballero, L., Martínez Gomez, E.: Comparison of NN. applied to time series prediction. *Prospectiva* **13**(2), 88 (2015)
19. Sartori, M.A., Antsaklis, P.J.: A simple method for deriving bounds on the size and training of MLP NN. *IEEE Trans. NN* **2**(4), 467–471 (1991)
20. Tamura, S., Tateishi, M.: Capacity of a four-layer feedforward NN.: four layers versus three. *IEEE Trans. NN* **8**(2), 251–255 (1997)
21. Jolliffe, I.T.: Principal component analysis and factor analysis. In: *Principal Component Analysis*, vol. 2, pp. 115–128 (1986)
22. Jolliffe, I.T.: Generaliz. of PCA using iterative data scaling. *J. Roy. Stat. Soc. Ser. C.* **21**(3), 184–195 (1972)
23. Grané, A.: Principal component analysis. Technical report, Universidad Carlos III, Department of Statistics, Madrid (2002)
24. Scikit-yb.org. Elbow method - yellow documentation v1.5. 2019. <https://www.scikit-yb.org/en/latest/api/cluster/elbow.html>. Accessed 10 Nov 2022
25. Stanford University. Sum of squared errors. [https://hlab.stanford.edu/brian/error\\_sum\\_of\\_squares.html](https://hlab.stanford.edu/brian/error_sum_of_squares.html). Accessed 10 Nov 2022
26. Kleine, D.: Detection of knee/elbow points in a graph (2021). <https://towardsdatascience.com/detecting-knee-elbow-points-in-a-graphd13fc517a63c>. Accessed 14 Nov 2022
27. Chng, Z.M.: Using Activation Functions in Neural Networks (2022). <https://machinelearningmastery.com/using-activation-functions-in-neural-networks/>. Accessed 01 May 2023
28. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. *Nature* **521**(7553), 436–444 (2016)
29. Sharma, S., Sharma, S., Athaiya, A.: Activation functions in neural networks. *Int. J. Eng. Appl. Sci. Technol.* **04**(12), 310–316 (2020)
30. Feng, J., Lu, S.: Performance analysis of various activation functions in ANN. *J. Phys. Conf. Ser.* **1237**(2) (2019)
31. Trevisan, V.: Comparing Robustness of MAE, MSE and RMSE. *Towards Data Science* (2022). <https://towardsdatascience.com/comparing-robustness-of-mae-mse-and-rmse-6d69da870828>. Accessed 29 Apr 2023



# Comparing How Python and R Estimate Granger-Causality in the Frequency Domain

Matteo Farne<sup>1</sup>  and Meng Yang<sup>2</sup> 

<sup>1</sup> Dipartimento di Scienze Statistiche, University of Bologna, Bologna 40126, Italy  
[matteo.farne@unibo.it](mailto:matteo.farne@unibo.it)

<sup>2</sup> Davis, Department of Statistics, University of California, Davis, USA  
[meyang@ucdavis.edu](mailto:meyang@ucdavis.edu)

<https://www.unibo.it/sitoweb/matteo.farne>

**Abstract.** This paper deals with the estimation of unconditional and conditional Granger-causality spectrum in the frequency domain. We describe two Python routines that parallel the existing R routines in computing these two quantities via package *grangers*. We present a simulation study showing that under zero-causality processes Python routines tend to perform slightly better than R, while under low-causality processes R routines perform quite better, because Python is less sensitive than R to small causality parameters. This difference can be attributed to the intrinsic VAR order selection procedure of the two packages.

**Keywords:** Granger-causality spectrum · VAR model · Python

## 1 Introduction

This paper deals with the estimation of Granger-causality (GC) spectrum in the frequency domain (FD). As explained in [7], GC can be a decisive tool to identify temporal cycles in the causal relationship from one time series to another one. The strength of the single cycle period  $1/\omega$  can be inferred by the strength of the corresponding GC spectrum at frequency  $\omega$ , where  $\omega \in [-\pi, \pi]$ . The relationship strength can be assessed both unconditionally, and conditionally on a third time series. For these purposes, [5, 10] and [6, 11] defined the unconditional and the conditional GC spectrum in the frequency domain, respectively. We refer to Section 2 in [7] for their definitions.

Granger-causality has been massively explored in recent years. The most analyzed topics by GC in the FD include environmental science [3], macroeconomics [16], and neuroscience [4]. In [7], a bootstrap test for both types of GC in the FD has been proposed, which complements the previous parametric test of [2], implemented in STATA by [13]. Some other relevant nonparametric inferential approaches for GC have been proposed in [8, 9, 14]. A relevant alternative strategy for unconditional and conditional GC, implemented in MATLAB, is the



MVGC toolbox [1], which described an exhaustive computational and inferential framework based on the exploitation of the estimated auto-covariance sequence. A very recent development suggests a new estimation and testing method based on a kernel approach [15].

In this paper, we explore the possibility to estimate both unconditional and conditional GC spectra via the original methods of [5] and [6] in Python, in comparison with the existing computational routines in R (see package *grangers* at the link <https://CRAN.R-project.org/package=grangers>). In particular, in this context GC estimation in the FD can be read as an indirect estimation problem: first, a bivariate or trivariate VAR model (see [12]) must be estimated, for unconditional or conditional GC respectively, and second, GC must be computed starting from the estimated VAR coefficient matrices. Such procedure is clearly affected by the VAR model estimation in the first step. For this reason, our aim is to measure the difference in terms of GC estimation between R and Python, which indirectly points to the difference in VAR model estimation by the two softwares.

The paper proceeds as follows. Section 2 presents the two new Python routines and compares their construction to the corresponding R routines. Section 3 contains a wide simulation study, where several scenarios are tested for both unconditional and conditional GC estimation. Therein, we report several performance metrics showing that Python and R behave differently in the presence of zero-causality and low-causality processes. In the end, Sect. 4 provides some concluding remarks.

## 2 Granger-Causality in the Frequency Domain

In this section, we show how the two Python functions computing respectively unconditional and conditional Granger causality spectra, named `granger_unconditional` and `granger_conditional`, are built, and we perform a systematic comparison with the two corresponding functions from the R package *grangers*, namely, *Granger.unconditional* and *Granger.conditional*.

### 2.1 Estimating Unconditional GC Spectrum

Python routines require several packages: *pandas*, *numpy*, *scipy*, *matplotlib*, *cmath*, *math*, *time*, *warnings*, *statsmodels.api* and *statsmodels.tsa*, that contains the *VAR* package. In the following, we describe the parameters of `granger_unconditional` function:

- *x*: the effect time series;
- *y*: the cause time series; *x* and *y* must be integer or float objects with the same length; otherwise, an error message will be displayed;
- *maxlag*: the maximum number of lags to check for order selection. It defaults to  $3 \times (T/100)$ . It must be integer or none (see `select_order` function from *VAR* package);

- *ic\_chosen*: the chosen information criterion for VAR order selection. It can be {'aic', 'fpe', 'hqic', 'bic', 'None'}. In more detail, the information criterion to use for VAR order selection can be Akaike ('aic'), final prediction error ('fpe'), Hannan-Quinn ('hqic') and Bayesian a.k.a. Schwarz ('bic');
- *type\_chosen*: trend type in VAR estimation. It can be {'c', 'ct', 'ctt', 'n'}, that is: 'c' (constant), 'ct' (constant and trend), 'ctt' (constant, linear and quadratic trend), 'n' (no constant, no trend);
- *plot*: it can be {'T', 'F'}, i.e. 'T'(True) or 'F'(False).

The results are stored in a *GrangerResult* class object, containing the following components:

- *frequency*: the Fourier frequencies on which the spectrum is computed, i.e.,  $\pi f/T$ ,  $f = 1, 2, \dots, T/2$ ;
- *n*: the time series length  $len(x)$ ;
- *Unconditional\_causality\_y.to.x*: estimated Granger-causality unconditional spectrum from  $y$  to  $x$ ;
- *roots*: estimated VAR (with  $x$  and  $y$ ) roots;
- *Lag Order*: estimated VAR lag order.

Referring to Section 2.1 in [7], the calculation of GC spectrum in the frequency domain,  $h_{Y \rightarrow X}(\omega)$ ,  $\omega \in [-\pi, \pi]$ , requires the computation of the transfer function  $\mathbf{P}(\omega)$  from the  $2 \times 2$  coefficient matrices  $\mathbf{A}_j$ ,  $j = 1, \dots, p$ , of the estimated VAR with  $x$  and  $y$ , the residual covariance matrix  $\mathbf{S}$ , and the transformed transfer function matrix  $\tilde{\mathbf{P}}(\omega)$ .

## 2.2 Estimating Conditional GC Spectrum

Compared to `granger_unconditional` function, `granger_conditional` function requires the same input arguments, with one important addition, namely, the conditioning time series  $z$ , which must be as long as  $x$  and  $y$ ; otherwise, an error message will be displayed. The function output is a *GrangerResult* object similar to the one returned by `granger_unconditional`, containing the following components:

- *frequency*: the Fourier frequencies on which the spectrum is computed, i.e.,  $h/T$ ,  $h = 1, 2, \dots, T$ ;
- *n*: the time series length  $len(x)$ ;
- *Conditional\_causality\_y.to.x.on.z*: estimated Granger-causality conditional spectrum from  $y$  to  $x$  given  $z$ ;
- *roots\_1*: estimated VAR roots ( $x$  and  $z$ );
- *roots\_2*: estimated VAR roots ( $x$ ,  $y$  and  $z$ );
- *Lag Order\_1*: estimated VAR lag order ( $x$  and  $z$ );
- *Lag Order\_2*: estimated VAR lag order ( $x$ ,  $y$  and  $z$ ).

Referring to Section 2.1 in [7], the calculation of  $h_{Y \rightarrow X|W}(\omega)$  requires the transfer function  $\mathbf{G}(\omega)$  of the VAR with  $x$  and  $z$ , the transfer function  $\mathbf{P}'(\omega)$  of the VAR with  $x$ ,  $y$ , and  $z$ , the transformed transfer function  $\mathbf{Q}(\omega)$ , and the residual covariance matrix of the VAR with  $x$ ,  $y$ , and  $z$ ,  $\Sigma_3$ .

### 2.3 Coding Differences Between R and Python

Both `granger_unconditional` and `granger_conditional` require many array definitions, exactly as the corresponding R functions `granger.unconditional` and `granger.conditional`. The main differences between the two softwares regard array indexing: Python indexes at 0 instead of 1 like R and other programming languages. Then, the most important difference is that `VAR` function in Python can return 0 as the best VAR order, while the corresponding `VAR` function in R does not. As a consequence, this can easily lead to error messages in Python, because a `VAR(0)` model returns a constant GC spectrum. This typically happens under zero or low causality processes, as we will see in the next section. Therefore, the fundamental Python back-end functions had to be adjusted compared to R. We just added “ifelse” clause to help locate the correct coefficient matrices when available and save them correctly.

## 3 Simulation Study

### 3.1 Settings

Suppose that the stochastic processes  $X_t$  and  $Y_t$ , jointly covariance-stationary, follow a non-singular  $VAR(p)$  model. Defining the bivariate stochastic process  $\mathbf{Z}_t = [X_t, Y_t]'$ , we have

$$\mathbf{Z}_t = \mathbf{A}_1 \mathbf{Z}_{t-1} + \dots + \mathbf{A}_p \mathbf{Z}_{t-p} + \epsilon_t, \quad (1)$$

where  $\epsilon_t$  is a zero-mean bivariate stochastic process with covariance matrix  $\mathbf{\Sigma}_2$  and  $\mathbf{A}_1, \dots, \mathbf{A}_p$  are  $2 \times 2$  coefficient matrices. We stress that  $\epsilon_t$  has no auto-covariance structure at any lag different from 0.

We simulate  $R = 300$  replicates from a VAR process in the form (1) with different features. The process length is fixed to  $T = 300$ . We set  $\mathbf{\Sigma}_2 = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ . We apply to procedure to estimate model (1): function `var` in R and function `var_model` in Python. The two procedures mainly differ by the selection of the VAR lag  $p$ , as explained in Sect. 2. We select  $p$  both by BIC and AIC criterion, and we both set no trend and a constant trend in (1). Then, we estimate the Granger-causality (GC) unconditional spectrum  $\hat{h}_{Y \rightarrow X}(\omega)$ , where the coefficient matrices  $\mathbf{A}_j, j = 1, \dots, p$ , and the error covariance matrix  $\mathbf{\Sigma}_2$  are replaced by the corresponding SURE estimates [17].

We consider the following scenarios featuring different unconditional GC shapes. The first three scenarios aim at testing zero-causality processes, where  $\mathbf{A}_{k,(j_1 j_2)} = 0, j_1 \neq j_2, k = 1, 2, \dots$ , also including the possible impact of residual correlation  $\rho$  and autoregressive coefficients  $\mathbf{A}_{1,(jj)}, j = 1, 2$ :

- scenario 1:  $p = 0, \rho = 0$ , with  $h_{Y \rightarrow X}(\omega) = 0, \omega \in [-\pi, \pi]$ ;
- scenario 2:  $p = 0, \rho = 0.5$ , with  $h_{Y \rightarrow X}(\omega) = 0, \omega \in [-\pi, \pi]$ ;
- scenario 3:  $p = 1, \rho = 0, \mathbf{A}_{1,(jj)} = 0.5, j = 1, 2$ , with  $h_{Y \rightarrow X}(\omega) = 0, \omega \in [-\pi, \pi]$ .

Then, we consider three VAR(1) processes with non-zero causality at all frequencies and a reverse S-shape, with the aim to test the impact of the causality coefficient  $\mathbf{A}_{1,(j2)}$ ,  $j = 1, 2$ , onto the GC estimation process:

- scenario 4:  $p = 1$ ,  $\mathbf{A}_{1,(j2)} = 0.2$ ,  $j = 1, 2$ ;
- scenario 5:  $p = 1$ ,  $\mathbf{A}_{1,(j2)} = 0.5$ ,  $j = 1, 2$ ;
- scenario 6:  $p = 1$ ,  $\mathbf{A}_{1,(j2)} = 0.8$ ,  $j = 1, 2$ .

We now deal with the conditional case. First, we simulate  $R = 300$  replicates from a VAR in the form (1) on  $[X_t, W_t]'$ , with covariance matrix of the noise terms  $\Sigma_{2'} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$  and, subsequently, we simulate  $R = 300$  replicates from a VAR in the form (1) on  $[X_t, Y_t, W_t]'$ , where, however, the coefficient matrices  $\mathbf{A}_k$ ,  $k = 1, \dots, p$ , are  $3 \times 3$ , and the covariance matrix of the noise terms is

$$\Sigma_{\mathbf{3}} = \begin{bmatrix} 1 & \rho_{XY} & 0 \\ \rho_{XY} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then, we estimate the Granger-causality (GC) conditional spectrum  $\widehat{h}_{Y \rightarrow X|W}(\omega)$ , where the coefficient matrices and the error covariance matrix are replaced by the corresponding SURE estimates.

We consider the following conditional GC scenarios, proceeding in analogous way *wrt* the unconditional case. First, we consider the following zero-causality processes, varying the residual correlation parameter  $\rho_{XY}$  and the autoregressive parameters  $\mathbf{A}_{1,(jj)}$ ,  $j = 1, 2$ :

- scenario 1:  $p = 0$ ,  $\rho_{XY} = 0$ ,  $h_{Y \rightarrow X|W}(\omega) = 0$ ,  $\omega \in [-\pi, \pi]$ ;
- scenario 2:  $p = 0$ ,  $\rho_{XY} = 0.5$ ,  $h_{Y \rightarrow X|W}(\omega) = 0$ ,  $\omega \in [-\pi, \pi]$ ;
- scenario 3:  $p = 1$ ,  $\rho_{XY} = 0$ ,  $\mathbf{A}_{1,(jj)} = 0.5$ ,  $j = 1, 2$ ,  $h_{Y \rightarrow X|W}(\omega) = 0$ ,  $\omega \in [-\pi, \pi]$ .

Then, we set  $p = 1$  and we test the impact of the causality parameter from  $Y_t$  to  $X_t$ , via the following scenarios with a reverse S-shape:

- scenario 4:  $p = 1$ ,  $\mathbf{A}_{1,(j2)} = 0.2$ ,  $j = 1, 2$ ;
- scenario 5:  $p = 1$ ,  $\mathbf{A}_{1,(j2)} = 0.5$ ,  $j = 1, 2$ ;
- scenario 6:  $p = 1$ ,  $\mathbf{A}_{1,(j2)} = 0.8$ ,  $j = 1, 2$ .

### 3.2 Metrics

We desire to measure the performance of R and Python in estimating unconditional and conditional GC in the frequency domain over the Fourier frequencies  $\omega_f = \pi f/T$ ,  $f = 1, \dots, F$ . We set  $F = T/2$ . We obtain the estimated spectra  $\widehat{h}_{Y \rightarrow X}^{(r)}(\omega_f)$  and  $\widehat{h}_{Y \rightarrow X|W}^{(r)}(\omega_f)$ ,  $f = 1, \dots, F$ ,  $r = 1, \dots, R$ ,  $R = 300$ . We explicit the  $F$ -dimensional vectors  $\widehat{\mathbf{h}}_{Y \rightarrow X}^{(r)}$  and  $\widehat{\mathbf{h}}_{Y \rightarrow X|W}^{(r)}$ , with  $r = 1, \dots, R$ , to be compared to the true  $\mathbf{h}_{Y \rightarrow X}^{(r)}$  and  $\mathbf{h}_{Y \rightarrow X|W}^{(r)}$ , respectively. We also extract the selected VAR lag order  $p^{(r)}$ ,  $r = 1, \dots, R$ . Then, we derive the following performance measures:

- the selected VAR lag order error  $p_{err} = R^{-1} \sum_{r=1}^R \mathbb{1}(p^{(r)} \neq p)$ ;
- the overall root mean squared error  $err_M^{(U)} = R^{-1} \sum_{r=1}^R \|\widehat{\mathbf{h}}_{Y \rightarrow X}^{(r)} - \mathbf{h}_{Y \rightarrow X}^{(r)}\|$  and  $err_M^{(C)} = R^{-1} \sum_{r=1}^R \|\widehat{\mathbf{h}}_{Y \rightarrow X|W}^{(r)} - \mathbf{h}_{Y \rightarrow X|W}^{(r)}\|$ ;
- the frequency-wise root mean square errors  $err_M^{(U)}(f) = R^{-1} \sum_{r=1}^R \|\widehat{\mathbf{h}}_{Y \rightarrow X}^{(r)}(\omega_f) - \mathbf{h}_{Y \rightarrow X}^{(r)}(\omega_f)\|$  and  $err_M^{(C)}(f) = R^{-1} \sum_{r=1}^R \|\widehat{\mathbf{h}}_{Y \rightarrow X|W}^{(r)}(\omega_f) - \mathbf{h}_{Y \rightarrow X|W}^{(r)}(\omega_f)\|$ , with  $f = 1, \dots, F$ .

### 3.3 Results

Table 1 reports the performance of VAR lag order selection by BIC and AIC for Python and R. We can note that, as expected, AIC systematically reports a larger error than BIC criterion. Comparing Python and R, we can note that under scenarios 3, 5 and 6 the performance is identical. We identify relevant differences under scenarios 1 and 2, where R is forced to estimate 1 lag, and scenario 4, where Python wrongly selects 0 lags 35% of times by BIC, and 13% of times by AIC. Under the same scenario, R is never wrong by BIC, and is equally wrong by AIC. This means that, under zero-causality processes, R cannot be able to select zero lags, while under low-causality processes, Python selects zero lags quite often, unlike R.

**Table 1.** Selected VAR lag order error  $p_{err}$  for R and Python under the six considered unconditional scenarios where the VAR order is selected by BIC and AIC criterion and a constant trend or no trend is added.

Scenario	constant				none			
	BIC		AIC		BIC		AIC	
	Python	R	Python	R	Python	R	Python	R
	$p_{err}$	$p_{err}$	$p_{err}$	$p_{err}$	$p_{err}$	$p_{err}$	$p_{err}$	$p_{err}$
1	0	1	0.13	1	0	1	0.13	1
2	0	1	0.12	1	0	1	0.12	1
3	0	0	0.1633	0.1633	0	0	0.1633	0.16
4	0.35	0	0.13	0.12	0.35	0	0.13	0.11
5	0	0	0.1033	0.1033	0	0	0.1033	0.0967
6	0	0	0.1433	0.1433	0	0	0.1433	0.1433

Table 2 reports the same metric for the six conditional scenarios considered, depicting a similar picture compared to Table 1. We can note that Python is now always wrong under scenario 4, because it is not able to identify the only lag under causality parameter 0.2. The same occurs almost always also under scenario 5 (causality parameter 0.5). This effect is a bit mitigated but still strong even by AIC. Under scenario 6, this effect disappears, but both softwares occasionally select more than 1 lag. On the other hand, R still cannot select lag 0 under scenarios 1 and 2 (zero-causality).

**Table 2.** Selected VAR lag order error  $p_{err}$  for R and Python under the six considered conditional scenarios where the VAR lag is selected by BIC and AIC criterion and a constant trend and no trend is added.

Scenario	constant				none			
	BIC		AIC		BIC		AIC	
	Python	R	Python	R	Python	R	Python	R
	$p_{err}$	$p_{err}$	$p_{err}$	$p_{err}$	$p_{err}$	$p_{err}$	$p_{err}$	$p_{err}$
1	0	1	0.02	1	0	1	0.02	1
2	0	1	0.0133	1	0	1	0.0133	1
3	0	0	0.1633	0.1633	0	0	0.1633	0.1567
4	1	0	0.9167	0.13	1	0	0.9167	0.1233
5	0.9767	0	0.6167	0.18	0.9767	0	0.6167	0.2
6	0.15	0.15	0.8467	0.8467	0.15	0.1633	0.8467	0.8867

Table 3 reports the root mean square error metrics for all unconditional scenarios. We can notice that Python is doing slightly better than R under zero-causality scenarios, while Python pays a price under scenario 4, which is larger than the gap observed under scenarios 1 and 2.

**Table 3.** Root mean square error  $err_M^{(U)}$  and its standard deviation  $err_S^{(U)}$  for R and Python under the six considered unconditional scenarios where the VAR lag is selected by BIC and AIC criterion and a constant trend is added.

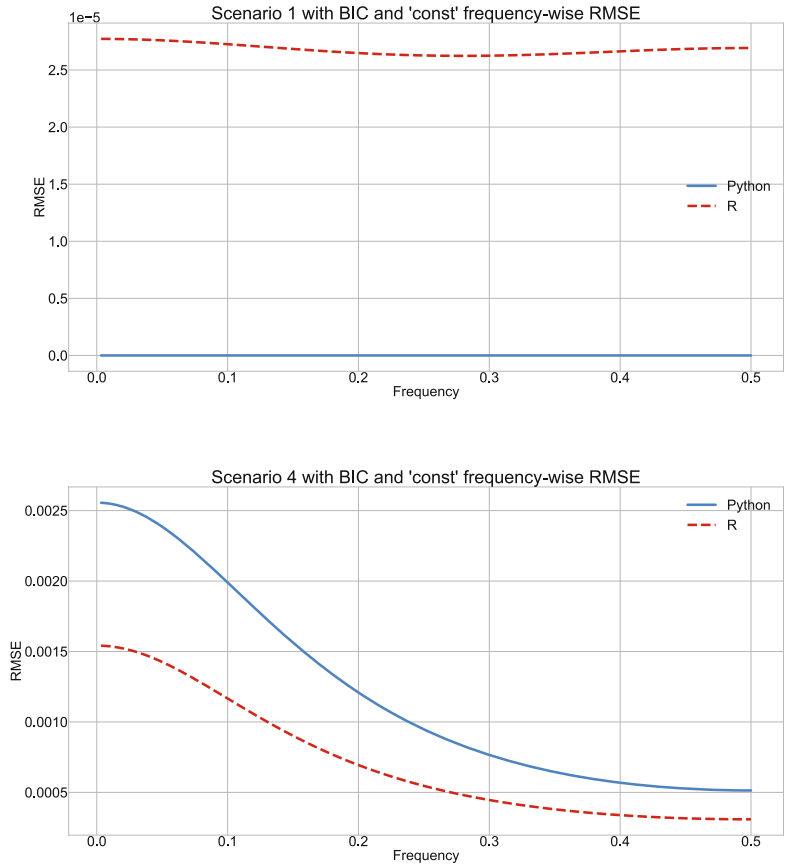
Scenario	BIC				AIC			
	Python		R		Python		R	
	$err_M^{(U)}$	$err_S^{(U)}$	$err_M^{(U)}$	$err_S^{(U)}$	$err_M^{(U)}$	$err_S^{(U)}$	$err_M^{(U)}$	$err_S^{(U)}$
1	0	0	0.0032	0.0041	0.0013	0.0053	0.0055	0.0097
2	0	0	0.0047	0.0066	0.0018	0.0078	0.0063	0.0096
3	0.0042	0.0063	0.0043	0.0063	0.0065	0.0096	0.0065	0.0096
4	0.0358	0.0242	0.0203	0.0172	0.0229	0.0196	0.0227	0.0193
5	0.3444	0.0378	0.3443	0.0377	0.3439	0.0406	0.3439	0.0406
6	1.2037	0.0957	1.2037	0.0957	1.2054	0.1003	1.2054	0.1003

Table 4 reports the same metrics for conditional scenarios. We can note that under scenarios 1 and 2 R is doing worse than Python, both by BIC and AIC, while under scenarios 4 and 5 Python is doing really worse than R when considering BIC criterion. When passing by a constant to none trend, the patterns of Tables 1 and 2 still hold.

Figure 1 reports the comparison of frequency-wise mean square error for R and Python for unconditional scenarios 1 and 4. We can note that under zero causality, R has a small error while Python is never wrong. Under low causality, Python error is relevantly larger than R error, and decreases according to the reverse S-shape.

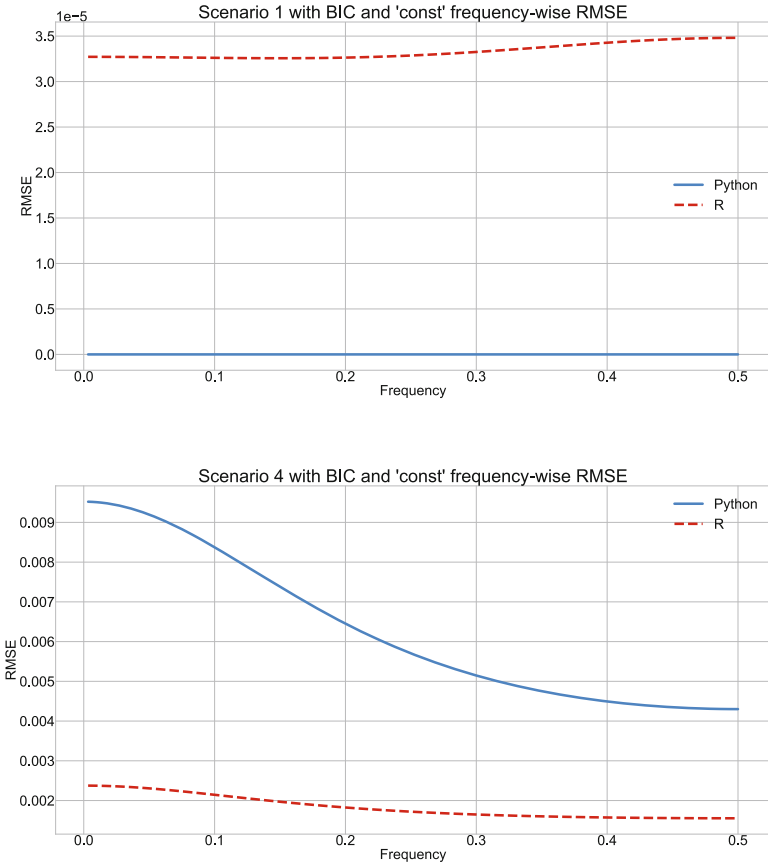
**Table 4.** Root mean square error  $err_M^{(C)}$  and its standard deviation  $err_S^{(C)}$  under the six considered conditional scenarios where the VAR lag is selected by BIC and AIC criterion and a constant trend is added.

Scenario	BIC				AIC			
	Python		R		Python		R	
	$err_M^{(C)}$	$err_S^{(C)}$	$err_M^{(C)}$	$err_S^{(C)}$	$err_M^{(C)}$	$err_S^{(C)}$	$err_M^{(C)}$	$err_S^{(C)}$
1	0	0	0.0031	0.0049	0.0001	0.0012	0.0050	0.0082
2	0	0	0.0044	0.0060	0.0002	0.0030	0.0063	0.0093
3	0.0043	0.0060	0.0043	0.0060	0.0062	0.0089	0.0062	0.0089
4	0.0976	0	0.0389	0.0178	0.0755	0.0128	0.0388	0.0177
5	0.4644	0.0469	0.1823	0.0430	0.3202	0.1520	0.1807	0.0441
6	0.3613	0.0568	0.3613	0.0568	0.3605	0.0567	0.3605	0.0567



**Fig. 1.** Comparing Python and R frequency-wise mean square error  $err_M^{(U)}(f)$  under unconditional scenarios 1 and 4.

Figure 2 reports the same comparison under conditional scenarios 1 and 4. The identifiable pattern is exactly the same.



**Fig. 2.** Comparing Python and R frequency-wise mean square error  $err_M^{(C)}(f)$  under conditional scenarios 1 and 4.

### 4 Conclusions

In this paper, we have described two Python routines for the estimation of unconditional and conditional Granger-causality spectra, and we have compared their performance to that of the existing corresponding routines in R. We have discovered that under zero-causality processes Python is doing better, because it is able to select 0 lags in the relative VAR process, while under low-causality processes, R is doing relevantly better, because Python tends to wrongly select 0 lags by BIC and (less frequently) by AIC, unlike R. Relevant future work includes testing these results against the MVGC MATLAB toolbox proposed in [1], and implementing in R and Python the new estimator proposed in [15].





## References

1. Barnett, L., Seth, A.-K.: The MVGC multivariate Granger-causality toolbox: a new approach to Granger-causal inference. *J. Neurosci. Methods* **223**, 50–68 (2014)
2. Breitung, J., Candelon, B.: Testing for short-and long-run causality: a frequency-domain approach. *J. Econom.* **132**(2), 363–378 (2006)
3. Candelon, B., Hasse, J.-B.: Testing for causality between climate policies and carbon emissions reduction. *Finance Res. Lett.* 103878 (2023)
4. Faes, L., et al.: A new framework for the time-and frequency-domain assessment of high-order interactions in networks of random processes. *IEEE Trans. Signal Process.* **70**, 5766–5777 (2022)
5. Geweke, J.-F.: Measurement of linear dependence and feedback between multiple time series. *J. Am. Stat. Assoc.* **77**, 304–313 (1982)
6. Geweke, J.-F.: Measures of conditional linear dependence and feedback between time series. *J. Am. Stat. Assoc.* **79**, 907–915 (1984)
7. Farnè, M., Montanari, A.: A bootstrap method to test Granger-causality in the frequency domain. *Comput. Econ.* **59**(3), 935–966 (2022)
8. Hidalgo, J.: Nonparametric test for causality with long-range dependence. *Econometrica* **68**, 1465–1490 (2000)
9. Hidalgo, J.: A bootstrap causality test for covariance stationary processes. *J. Econom.* **126**, 115–143 (2005)
10. Hosoya, Y.: The decomposition and measurement of the interdependency between second-order stationary processes. *Probab. Theory Relat. Fields* **88**(4), 429–444 (1991)
11. Hosoya, Y.: Elimination of third-series effect and defining partial measures of causality. *J. Time Ser. Anal.* **22**(5), 537–554 (2001)
12. Lütkepohl, H.: Vector autoregressive models. In: *Handbook of Research Methods and Applications in Empirical Macroeconomics*, pp. 139–164. Edward Elgar Publishing (2013)
13. Tastan, H.: Testing for spectral Granger causality. *Stand Genomic Sci.* **15**(4), 1157–1166 (2015)
14. Nishiyama, Y., Hitomi, K., Kawasaki, Y., Jeong, K.: A consistent nonparametric test for nonlinear causality-specification in time series regression. *J. Econom.* **165**(1), 112–127 (2011)
15. Taufemback, C.-G.: Non-parametric short-and long-run Granger causality testing in the frequency domain. *J. Time Ser. Anal.* **44**(1), 69–92 (2023)
16. Usman, K., Bashir, U.: The causal nexus between imports and economic growth in China, India and G7 countries: granger causality analysis in the frequency domain. *Heliyon* **8**(8), e10180 (2022)
17. Zellner, A.: An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Am. Stat. Assoc.* **57**(298), 348–368 (1962)



# Performance Analyses of AES and Blowfish Algorithms by Encrypting Files, Videos, and Images

Karrar Hamzah Mezher<sup>1</sup>  and Timur Inan<sup>2</sup> 

<sup>1</sup> Information Technology, Altinbas University, Istanbul, Turkey

Alkaabikarrar9@gmail.com

<sup>2</sup> Altinbas University, Istanbul, Turkey

Timur.inan@altinbas.edu.tr

**Abstract.** This study attempts to secure the stored images and restrict unauthorized individuals from accessing them to effectively protect the data of the various systems, whether it be an image, a video, or a text file. For the data encryption process, symmetric encryption algorithms have been proposed. The findings of two well-known encryption algorithms were compared to make sure the best methods were being employed. Given that both Blowfish and AES symmetric encryption use block ciphers with a lot of data, they were chosen. Both techniques are capable of encoding high-resolution facial images, and the encoded files can be analyzed using quantitative metrics like histograms and time elapsed with volume scaling. According to the results obtained using the suggested criteria, AES is preferred since it produces specific results in terms of picture and file encoding quality and accuracy, processing and execution speed, coding complexity, coding efficiency, and homogeneity. In conclusion, symmetric encryption methods protect the face recognition system faster and more effectively than alternative options. The AES algorithm is favored over others because of its enormous block space, encryption accuracy, and speed of implementation.

**Keywords:** AES · Blowfish · Symmetric Encryption · Histogram · Elapsed time

## 1 Introduction

This section will concentrate on a review of relevant literature on cryptography technology. Encryption technology has been employed in many organizations and security systems in modern technology to prevent unauthorized persons from entering as well as secure and protect various files. The data files contain a collection of information that may be used to determine a person's identity, special procedures, and technology; it is the most unique and extensively utilized key to reliability. Encryption technology works on the idea of taking a clear image, video, or file and encrypting it to protect the data contained within. It can be asserted that encryption techniques have priority in preventing hackers from accessing the content of the data. This paper will make a comparison of two algorithms to ensure the reliability of securing data, two highly capable encryption

algorithms are selected from symmetric cryptographic techniques. The content of the stored images, files, and videos is encrypted by both techniques. After comparing the results according to the processing standards, we get the best algorithm to secure the data stored.

The AES algorithm and Blowfish have been proposed as symmetric ciphers. Once the contents of the files have been encoded, the ciphered files and photographs are evaluated by mathematical criteria (Elapsed Time and Histogram.). In order to select the fastest algorithm, the average amount of time needed for the encryption process is computed. Based on the outcomes of the evaluation, the images and data that have been stored are encoded using the best and most effective algorithm.

## 2 Related Work

In this paper, the researchers paid attention to encryption algorithms to protect different files, including the AES and DES algorithms, and they reached several important features of the AES algorithm and presents some of the research conducted on it, to evaluate the performance of AES for data encryption under different parameters. AES The research's findings demonstrate that AES can offer greater security compared to other algorithms like DES, 3DES, etc. The authors claim that the AES method, one of the most popular and widely used symmetric block ciphers, contains a proprietary algorithm with a unique structure for encrypting and decrypting sensitive data. There is no proof that it can be cracked, despite being used in hardware and software across the globe. Each of the three key sizes that AES can handle—AES 128, 192, and 256 bits—may be used to encrypt and decrypt various kinds of data [1].

Khatri–Valmik, and Kshirsagar, they conclude that data security requirements today are higher than they were in the past. They suggest several cryptographic methods, such as the Blowfish algorithm derived from the Feistel network block cipher, which has a configurable key size that can be up to 448 bits and a block size of 64 bits. They concluded that the Blowfish algorithm is not protected by a patent and can be utilized by anybody, anywhere. Instead of the Data Encryption Standard (DES) or the International Data Encryption Algorithm (IDEA), the researchers utilized the Blowfish algorithm [2]. It requires a variable-length key, making it ideal for both domestic and international use. The findings showed that, in contrast to the other algorithms available, the Blowfish technique is not constrained by patents, is free to use by anybody in any circumstance, is adaptable, and offers a high level of data security [3].

Devi, A, et al., discuss how network security, which is in charge of protecting information exchanged between networked computers, is the most crucial element of information security. This paper discusses the encryption and decryption of images using the DES, AES, and Blowfish algorithms. Where academics in the modern world have reached the challenge of sending an image across the Internet from one network to another without the potential of penetration. Appropriate encryption can prevent access to the sent images' content, and decryption must be used to prevent illegal access. As a result, the researchers reviewed relevant studies, pointed out certain issues, and offered some conclusions that might be helpful for image coding. Several image encryption and decryption methods have been examined. The analysis and problem formulation were

based on their study. The authors of this study compare the DES, AES, and Blowfish algorithms to speed up image decryption and enhance performance. Future work will compare and analyze popular encryption techniques like AES and Blowfish experiment with images, files, and videos, and concentrate on reducing encoding and decoding times [4].

The Blowfish and AES algorithms' strengths and shortcomings have been investigated. Their research served as the foundation for problem analysis and formulation. The authors of this paper compare the AES and Blowfish algorithms for picture decoding speed and performance. The researchers determined that the AES algorithm is more resistant to hacking since it uses larger key sizes for encryption, such as 128, 192, and 256 bits. Furthermore, no one can hack into personal information. Around 2128 attempts are necessary to crack the technique's key lengths. Because it is a very secure protocol, it is extremely difficult to hack [5] for vulnerabilities, AES per block is always encrypted in the same way and is difficult to implement with software. As for the strengths and weaknesses of the Blowfish algorithm, the algorithm is one of the fastest block ciphers in public use and is freely available for anyone to use. The Blowfish algorithm's drawbacks are that it must receive a user's key from a distance rather than through a specific insecure transmission channel and that it performs poorly when compared to other algorithms in terms of serial throughput and time consumption during decryption [6].

### 3 Methodology

By giving a more thorough comparison of the algorithms, this research study investigates the various approaches used to encode files, photos, and videos to determine the most effective strategies [7]. Through the use of specialized techniques and encryption keys that are learned by both the transmitter and the recipient, the chaotic approach of scattering the data's contents during encryption is used. To recover the original data content, the encryption process is reversed during decryption [8]. The asymmetric encryption method is used over one secret key. The AES and Blowfish algorithm is the most recent and efficient symmetric encryption algorithm that is proposed in this work [9]. Some special criteria were used to evaluate the efficiency of encryption, such as calculating the elapsed time and the histogram technique, as well as measuring the size of files before and after encryption to find out the effect of the algorithm on the size of files, images and videos.

#### 3.1 Blowfish Algorithm

Symmetric box with suitable encryption and data protection is Blowfish. It is ideal for data security since it employs keys with varying lengths that range from 32 bits to 448 bits [10]. The Blowfish algorithm is a Feistel network with a symmetric key square shape. The square size is 64 bits, demonstrating evident encryption work in numerous instances. It has an extremely high encryption rate. Regardless of how great a presentation is required before any encryption can take place; genuine data encryption is very strong on huge microchips [3]. Bruce Schneier created Blowfish in 1993 [11]. The same secret key is used for both encryption and decryption. The block length of Blowfish is 64 bits.

It can employ keys with lengths ranging from 32 to 448 bits (128 bits is the norm). One primary key will generate 18 sub-keys. It takes 521 iterations to get the required number of subkeys. It has 16 rounds, is a Feistel cipher, and employs a big key based on S-boxes. CAST-128, which frequently employs a fixed number of S-boxes, occurs in the build [12]. The pseudo code of algorithm explained below [13, 14].

The input is a 64-bit data element,

x. Divide x into two 32-bit halves: xL, xR.

Then, for  $i = 1$  to 16:  $xL = xL \text{ XOR } P_i$   $xR = F(xL) \text{ XOR } xR$  Swap xL and xR

After the sixteenth round, swap xL and xR again to undo the last swap.

Then,  $xR = xR \text{ XOR } P_{17}$  and  $xL = xL \text{ XOR } P_{18}$ .

recombine xL and xR to get the ciphertext.

The 32-bit input is divided into four eight-bit quarters by the F-function, which then feeds the quarters into the S-boxes. The final 32-bit output is created by XORing the outputs together and adding them modulo 232.

### 3.2 AES (Advanced Encryption Standard)

It utilizes symmetric block cipher technology and is one of the most extensively used algorithms in the world [15]. It was produced in 2000 by the National Institute of Standards and Technology (NIST) in response to the need for new cryptographic technologies to perform in place of existing ones with security flaws [16]. AES has three key sizes: 128, 192, and 256 bits, with a 128-bit block size. AES performs multiple rounds of data conversion for each block. The key length is 10 for a key value of 128, 12 for a key value of 196, and 14 for a key value of 256. The block is turned into a 4\*4 matrix at the start of the encryption procedure, and then the cycles are executed. AES offers procedures for clear data content hashing, alternate bytes, row modification, mixed columns, and circular key addition in each round. The technique and steps are reversed during decoding [5]. Pseudo code for the AES algorithm is mentioned below [17, 18].

```



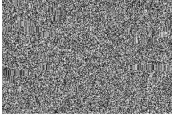

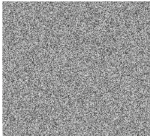
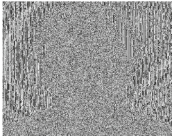






INPUT STATE, KEY
OUTPUT CIPHER
State=AddRoundKey(STATE, KEY[0,...,3] );
FOR i=1 to Round
{
    STATE = SubstituteBytes (STATE);
    STATE = ShiftRows (STATE)
    IF (i < Round)
        STATE = MixColumns (STATE)
    STATE =AddRoundKEY (STATE, KEY [ 4*i, ..., 4*i+3]);
}
CIPHER=STATE

```

### 4 Experimental Result

A text file, a video, and two high-resolution grayscale photographs of the face with varying proportions were used. The first image has set dimensions of  $300 \times 168$  pixels (as a small image), while the second image has changeable dimensions of  $3000 \times 3192$  pixels (as a large image). The AES and Blowfish algorithms are used to encrypt images, videos, and text files. The performance of picture encoding and decoding methods was evaluated using the criteria proposed in the simulation procedure. To obtain high-resolution photos, face photographs were sampled online from the face collection. The Python programming language is used to code all algorithms and simulations. To evaluate the efficiency of the algorithms and choose the best among them to achieve reliability and the greatest levels of security in data protection, elapsed time and histogram criteria for images are computed, as well as verifying the size of encrypted text and video files.

**Table 1.** Natural Shapes and Sizes of Original and Encrypted Images, Files and Videos.


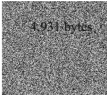

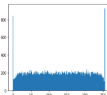
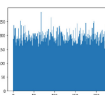




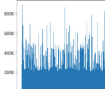
Process	Image	Original Image	AES	Blowfish
Encryption	Small Image (300 x 168)	 4,931 bytes	 28,692 bytes	 48,323 bytes
	Large Image (3000 x 3192)	 1,134,335 bytes	 5,321,223 bytes	 5,566,037 bytes
Encryption for Text File	 Text File 5,527 bytes	 5,552 bytes	 5,527 bytes	
Encryption for Video	 38,247 bytes	 38,272 bytes	 38,222 bytes	

Depending on Table 1, in image encryption, we notice that the AES algorithm is superior to the Blowfish algorithm for both images and that the size of the encrypted images is less in the AES algorithm. As for the video and image encoding, the size was very similar for both algorithms and thus the AES algorithm can be preferred in image encoding over the Blowfish algorithm.

### 4.1 Histogram Analysis

The image’s mean, maximum, and heterogeneity values were determined using histogram analysis. The proportion of data points that fall within a given range is shown in the histogram, which represents numerical data [19]. The graph shows the frequency distribution of the image’s pixel density values since the original images’ pixel density values are dispersed whereas encoded images are predicted to have uniform values [20]. According to Table 2, the histograms of the unencrypted and normal photos differ, avoiding data leakage.

**Table 2.** Histogram Graphics.

Process	Image	Original Image	AES	Blowfish	Histogram for AES	Histogram for Blowfish
Encryption	Small Image (300 x 168)	 4,931 bytes	 28,692 bytes	 48,323 bytes		
	Large Image (3000 x 3192)	 1,134,335 bytes	 5,321,223 bytes	 5,566,037 bytes		

Histograms of images encrypted using the AES and Blowfish algorithms are shown in Table 2. In contrast to Blowfish, which disperses the pixel values on the graph, when compared to the original photos and the encrypted images, where the pixel density values are uniform, the AES technique was successful in encrypting the image by evenly distributing the pixel density on the graph. Also, the table shows heterogeneity in the pixel values of the images encoded in the puffy network algorithm, as well as the lack of regulation in changing the content of the original images. Although the size of the encrypted images in the AES algorithm is less, it is more densely organized in changing the image content than the Blowfish algorithm, which greatly increases the size of the encrypted images.

### 4.2 The Elapsed Time

The operating times of encryption methods depend on the parameters of the chosen encryption method as well as on the amount of RAM, CPU, and disk space that is available [21]. The system becomes more dependable and secure when the algorithm does its tasks instantly. When evaluating the effectiveness of the algorithm, the speed of the encryption process is important [22].

**Table 3.** Elapsed time for AES and Blowfish Encryption Analyses.

Image	Encryption Time of AES	Encryption Time of Blowfish	File	Encryption Time of AES	Encryption Time of Blowfish
Small	0.0062	21.2011	Text File	0.0001	0.0018
Large	0.0001	57.6734	Video	0.0027	1.0000

The findings for both the AES and Blowfish algorithms' encryption operations per second are provided in Table 3. AES is faster than Blowfish in image encryption operations by a wide range. For text and video files, the encryption results were somewhat similar for each of the two algorithms, with a slight superiority for the AES algorithm.

## 5 Discussion

AES and Blowfish produced good results in the data encryption process while using histogram criteria, looking at the size of images and files, and measuring the amount of time that passed. The AES algorithm performs well in terms of processing speed, coding complexity, coding quality, and consistency. Blowfish encrypts data in 64-bit chunks, whereas AES encrypts data in 128-bit chunks. When compared to AES, Blowfish consumes less RAM. This study shows that the AES algorithm performs faster encryption of stored data of all types than the Blowfish approach.

## 6 Conclusion

We analyzed symmetric encryption techniques that make use of block cryptography, which is the most popular method for protecting huge data, to identify the best method for protecting data stored in any system. AES and Blowfish, the most recent, have both been employed. The AES method outperformed the Blowfish algorithm in several ways, including execution speed, results correctness, cryptographic quality and efficiency, and data size uniformity, according to a comparison utilizing quantitative and physical criteria. In terms of processing speed, precision, efficiency, and dependability, AES performs better than Blowfish. Additionally, using symmetric encryption algorithms to protect data systems is advised.

## References


1. Abdullah, A.M.: Advanced encryption standard (AES) algorithm to encrypt and decrypt data (2017)
2. Fernando, E., et al.: Performance comparison of symmetries encryption algorithm AES and DES with raspberry pi. In: 2019 International Conference on Sustainable Information Engineering and Technology (SIET), pp. 353–357 (2019)
3. Nehakhatri-Valmik, M., Kshirsagar, V.: Blowfish algorithm (2014)



4. Devi, A., Sharma, A., Rangra, A.: A review on DES, AES and blowfish for image encryption & decryption (2015)
5. Aleisa, N.: A Comparison of the 3DES and AES Encryption Standards (2015)
6. Nie, T., Song, C., Zhi, X.: Performance evaluation of DES and Blowfish algorithms. In: 2010 International Conference on Biomedical Engineering and Computer Science, pp. 1–4 (2010)
7. Kaur, M., Kumar, V.: A comprehensive review on image encryption techniques (2020)
8. Matta, P., Arora, M., Sharma, D.: A comparative survey on data encryption Techniques: Big data perspective (2021)
9. Dibas, H., Sabri, K.E.: A comprehensive performance empirical study of the symmetric algorithms: AES, 3DES, Blowfish and Twofish. In: 2021 International Conference on Information Technology (ICIT), pp. 344–349 (2021)
10. Timilsina, S., Gautam, S.: Analysis of Hybrid Cryptosystem Developed Using Blowfish and ECC with Different Key Size (2019)
11. Singh, P., Singh, K.: Image encryption and decryption using blowfish algorithm in MATLAB (2013)
12. Vengatesan, K., et al.: Secure data transmission through steganography with blowfish algorithm. In: Hemanth, D.J., Kumar, V.D.A., Malathi, S., Castillo, O., Patrut, B. (eds.) COMET 2019. LNDECT, vol. 35, pp. 568–575. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-32150-5\\_55](https://doi.org/10.1007/978-3-030-32150-5_55)
13. Christina, L., Joe Irudayaraj, V.: Optimized Blowfish encryption technique (2014)
14. Al-Neaimi, A.M.A., Hassan, R.F.: New Approach for Modifying Blowfish Algorithm by Using Multiple Keys (2011)
15. Chen, X.: Implementing AES encryption on programmable switches via scrambled lookup tables. In: Proceedings of the Workshop on Secure Programmable Network Infrastructure, pp. 8–14 (2020)
16. Batáry, P., et al.: The role of agri-environment schemes in conservation and environmental management (2015)
17. Lavanya, R., Karpagam, M.: Enhancing the security of AES through small scale confusion operations for data communication (2020)
18. Ortakci, Y., Abdullah, M.Y.: Performance analyses of AES and 3DES algorithms for encryption of satellite images. In: Innovations in Smart Cities Applications Volume 4: The Proceedings of the 5th International Conference on Smart City Applications, pp. 877–890 (2021)
19. Dhal, K.G., et al.: Histogram equalization variants as optimization problems: a review (2021)
20. Kar, C., Banerjee, S.: Tropical cyclones classification from satellite images using blocked local binary pattern and histogram analysis. In: Soft Computing Techniques and Applications: Proceeding of the International Conference on Computing and Communication (IC3 2020), pp. 399–407 (2021)
21. Gajaria, D., Adegbija, T.: Evaluating the performance and energy of STT-RAM caches for real-world wearable workloads (2022)
22. Dos Anjos, J.C., et al.: An algorithm to minimize energy consumption and elapsed time for IoT workloads in a hybrid architecture (2021)



# A Wireless Emergency Alerts System for Warning Disasters by Using Distributed Databases, GPS and Machine Learning Enabled API Services

Md. Abdullah Al Mamun<sup>1</sup>, Md. Tanvir Miah Shagar<sup>1</sup>, Meher Durdana Khan Raisa<sup>1</sup>,  
Md. Jubayer Hossain<sup>1</sup>, Utsa Chandra Sutradhar<sup>1</sup>, S. Rayhan Kabir<sup>2,3</sup> ,  
Anupam Hayath Chowdhury<sup>3</sup>, and Mohammad Kamrul Hasan<sup>2</sup>

<sup>1</sup> Department of Software Engineering, Daffodil International University, Dhaka, Bangladesh  
{abdullah35-3087, tanvir35-3088, meher35-2818, jubayer35-625, utsa35-523}@diu.edu.bd

<sup>2</sup> Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600  
UKM, Bangi, Selangor, Malaysia  
p126933@siswa.ukm.edu.my, mkhasan@ukm.edu.my

<sup>3</sup> Department of CSE, Asian University of Bangladesh, Dhaka, Bangladesh  
{rayhan923, anupamhc}@aub.edu.bd

**Abstract.** As most disasters are weather-related, location tracking is crucial during any event, and early warning systems can save lives. This article outlines the design and execution of a wireless emergency alerts model for catastrophic alerting based on data from distributed databases by using machine learning enabled Application Programming Interface (API) services. Using machine learning-based historical and current data from the NASA Open API and OpenWeatherMap API, this program is intended to forecast natural catastrophes and aid in their management. To forecast future disasters, the application analyses data from previous natural disasters, including floods, wildfires, tornadoes, cyclones, and hurricanes. Additionally, it gives users access to GPS (Global Positioning System) based evacuation routes and emergency shelter information. The purpose of this article is to assist individuals in the lead-up, during, and aftermath of a disaster by providing access to important resources and information. The software serves as a daily catastrophe management tool, notifying users of the most recent disasters and letting them know whether an incident is happening right now using machine learning-based data. This application can be an important tool to help individuals from disaster.

**Keywords:** Wireless Data Communication · Machine Learning · Disaster · GPS

## 1 Introduction

Catastrophic natural events are violent occurrences that mankind cannot prevent. As a result of natural causes, they can be dangerous and even deadly. The term “natural disaster” is used to describe any catastrophic event caused by the natural environment

that endangers human lives, property, or essential infrastructure. Extreme weather events, such as cyclones, have become more frequent and more powerful because of climate change, wreaking havoc on coastal towns [10]. Figure 1 shows the global natural disasters from 1970 to 2022 of wildfires, Extreme temperatures, extreme weather, and floods [12]. The risks associated with the climate and the potential for significant weather disasters are both exacerbated by global warming. Warmer air and water cause sea levels to rise, storms to intensify, and flooding to occur.

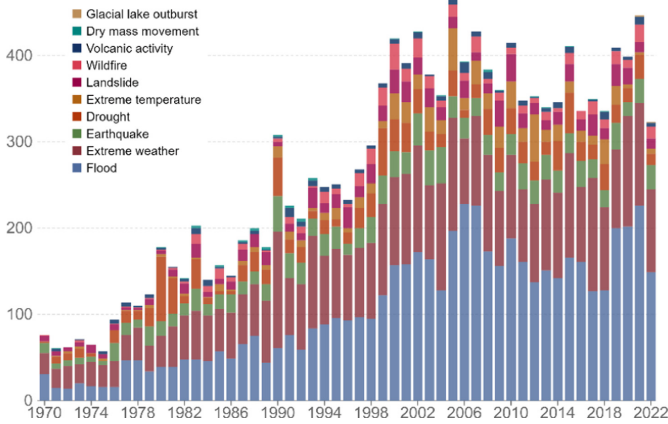


Fig. 1. Global natural disasters by type, 1970 to 2022 [12].

This portion briefly discusses the literature review or related works. Major destruction can be caused by natural disasters including floods, landslides, and earthquakes, often resulting in the loss of life and material possessions [5]. There is a need for early warning systems (EWS) that can identify and predict these calamities so that people can be prepared for them [2]. There is potential for improvements in catastrophe monitoring and EWS made possible by the implementation of cutting-edge technology like 3D-MEMS, NB-IoT, and deep learning neural networks [11]. In this study, researchers investigated how these tools might be applied to designing a service oriented IoT architecture for use in emergency management and weather forecasting [3]. Cloud of Things (CoT) and Blockchain based IoT (BLoT) are the advanced version of IoT (Internet of Things) where researchers focused on inflation [14, 15, 17]. Additionally, we'll look at how deep-learning neural networks may be used to predict the likelihood of flash floods in a certain area with pinpoint accuracy [4]. A recent study of wireless emergency alerts method has been focused on earthquakes in the West Coast, USA [1]. Since NASA Open and OpenWeatherMap two datasets contain real-time global data, after reviewing the previous literature, we realized that the disaster warning system should have a combination of these two datasets in addition to GPS so that a system can provide complete information, which is not available in previous studies. Our research has addressed this research gap.

Overall, Disaster Alert is a valuable contribution to research in disaster management and an essential tool for individuals facing natural catastrophes. Therefore, this paper provides the following contribution:

- The proposed Wireless Emergency Alert System is a machine learning system that uses weather data from NASA Open API and OpenWeatherMap API to predict natural disasters, is a significant contribution to the field of disaster management.
- In the proposed model, we have connected alerts both online and offline. Here, alerts can be sent automatically through GPS if people's phones are offline.
- By utilizing past and present real-time data, the app can forecast trends and patterns, helping individuals prepare and take necessary precautions.
- Additionally, the inclusion of GPS-based evacuation routes and information on local shelters both online and offline makes the app an essential tool for individuals during natural disasters.
- The integration of machine learning in the app's development opens new possibilities for improving weather predictions and ensuring people's safety.

In the introduction section we already discussed literature review, research gap, and contribution portions. Section 2 describes the proposed disaster alerting system and illustrates the model. Section 3 describes the results of the model. Finally, Sect. 4 shows the conclusion and future study of the research.

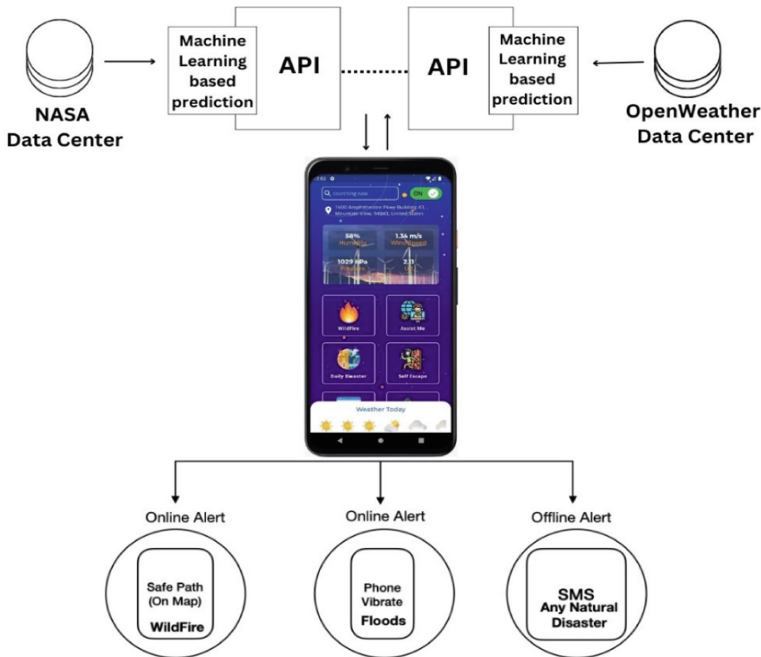
## 2 Wireless Emergency Alerts System for Warning Disasters

### 2.1 Machine Learning Enabled API Services

This proposed system serves us disaster alerts by collecting machine learning based predicted data from NASA Open API and OpenWeatherMap API so that people can take necessary steps against the disaster before it happens. It also shows weather, floods, wildfires, tornadoes, cyclones, and hurricanes. Python libraries of Short-term Ensemble Prediction have been as ensemble machine learning for classic weather forecasting in OpenWeatherMap API. This API data can deliver all the weather data required for decision-making for any location on the planet thanks to Convolutional Neural Network (CNN) solutions. In a study of ancient heritage buildings, researchers used CNN [18]. Machine learning is a widely used technique with which researchers have focused on pandemic stages identification [22], cyber-attack detection of smart grids [24] and IoT based sustainable cities in recent studies [25].

The Earth Science Data Systems (ESDS) Program at NASA is committed to using machine learning and is aware of its potential to significantly improve operations, expand the capabilities of current data systems, and make the most of Earth observation data. The Interagency Implementation and Advanced Concepts Team (IMPACT) is an interdisciplinary group that works to advance the ESDS objective of managing the Earth science data lifecycle to maximize the scientific value of NASA's missions and experiments for researchers, policymakers, and society. The IMPACT team of NASA, made up of experts in machine learning, computer science, and Earth science data, works to develop the tools and pipelines needed to apply machine learning algorithms to NASA Earth science datasets to enhance data discovery. Along with IMPACT, groups at the

Distributed Active Archive Centers (DAACs) and NASA Earth Observing System Data and Information System (EOSDIS) are using machine learning to analyze the data they distribute and archive. The ongoing work at NASA’s Goddard Earth Sciences Data and Information Services Center (GES DISC) to implement a machine learning framework using natural language processing (NLP) to make it simpler for GES DISC data users to find suitable datasets is one example. Distributed database based distributed machine learning have been focused on load forecasting process in smart grid network [15]. In this research, the application collected weather forecasting data from distributed data center through the APIs and the machine learning performed in the distributed data center sides. Furthermore, for security reasons, third-party data centers do not mention the specifics of the machine learning process in their documentation.



**Fig. 2.** Warning Disasters by using Distributed Databases, GPS and Machine Learning enabled API Services [Application Github Link: <https://github.com/Utsa05/Alert-App-Nasa>].

### 2.2 Proposed Mechanism

The app uses the NASA Open API and OpenWeatherMap API to collect weather-based data, which gives predictions for future disasters. Figure 2 shows how “Alert!” works. Usually, the predictions are sent as an alert through the app via notification and SMS. It also provides a safe route and shelter based on the map. Figure 2 shows the app model where the purpose of this app is to help people live safer lives in the face of natural disasters and calamities. The app offers services like safe zones, directions,

maps, real-time notifications, and alerts for disasters. It also has two special features, “*Daily Disaster*” and “*Self-Escape*,” that can also be used together. The Daily Disaster feature alerts users to danger zones and provides information on the cause of the danger, as well as a “*go-to-safe*” button that shows the shortest path to a safe zone on a map. The self-escape feature provides real-time notifications to users if they are in a dangerous position. Additionally, we have a feature called “*Wildfire*,” which shows all the previous and current wildfires on a Google map by using NASA JSON data [13]. The pseudocode of the application of proposed system as per described is shown in the algorithm below.

**Algorithm 1:** Pseudocode of Proposed Disaster Alerting System

```

1: BEGIN
2:   FUNCTION get_nasa_data():
3:     Retrieve Machine Learning based weather data from NASA API.
4:     //This function is set periodically to keep data up to date//
5:   END FUNCTION
6:   FUNCTION get_weather_data(latitude, longitude)
7:     Retrieve Machine Learning based data from OpenWeatherMAP API.
8:   END FUNCTION
9:   FUNCTION display_map():
10:    Display a map with data overlays by using get_weather_map_data from
    OpenWeatherMAP API
11:    Show Wildfire on map.
12:  END FUNCTION
13:  FUNCTION send_alerts(users, disaster_prediction):
14:    Sending alerts (Disaster Prediction Information) to users in the
    affected area by using their preferred method of contact (email,
    SMS, Push Notification, GPS)
15:  END FUNCTION
16:  WHILE (true): // Main Program: //
17:    Periodically call get_nasa_data(), call get_weather_data() , call
    get_weather_map_data()
18:    //This can be done once per day or more frequently as needed //
19:    IF (device is online)
20:      If the device is online, will get push notification alert for any
      disasters also. it will show in our app.
21:      //This can be done every hour or more frequently as needed//
22:    ELSE IF
23:      Else if device is offline, will get push notification Alert for
      any disasters.
24:      //This can be done every hour or more frequently as needed//
25:    END IF-ELSE
26:    Display_map()//Display map with data overlays//
27:    // User can exit the app or view alerts
28:    //Alerts are displayed as notifications or popups on the map//
29:  END WHILE
30: END

```

In this project, we used the prediction data obtained from the weather datasets of NASA Open and OpenWeatherMap datacenters using machine learning processes in the mentioned two distributed datacenters which were fed into our application via API. Additionally, for security reasons, third-party data centers do not specifically mention machine learning processes in their documentation. When internet service is not available, the importance of GPS in the proposed system is to concentrate more on disaster monitoring and management. This is due to the use of GPS to deliver highly accurate

real-time location information that can assist in managing every step of a disaster event, including pre-disaster, disaster, and post-disaster events.

### 3 Result and Finding Outcomes

The result section of this article shows the preliminary results of the application project where we illustrated some screenshots of the intelligent application. Research work on the validity of the model, data accuracy, and data analysis are in progress and may be published later. The natural disaster prediction and response app developed in this research project utilizes NASA data and Earth observations to provide accurate and up-to-date information to help people prepare for and respond to natural disasters. The app includes features such as Daily Disaster and Self Escape, which work together to provide users with information about danger zones and the shortest path to a safe zone. Additionally, the app includes a wildfire feature that uses NASA JSON data to show previous and current wildfires on a Google map.



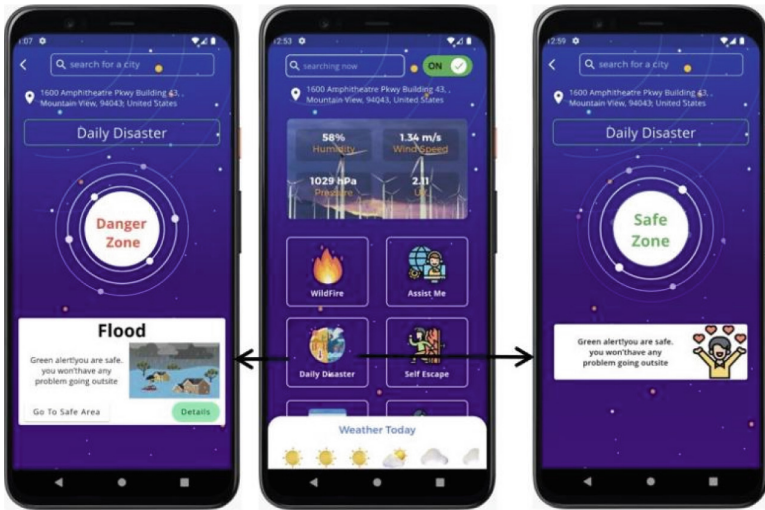
**Fig. 3.** Live Humidity, Wind Speed, Pressure & UV



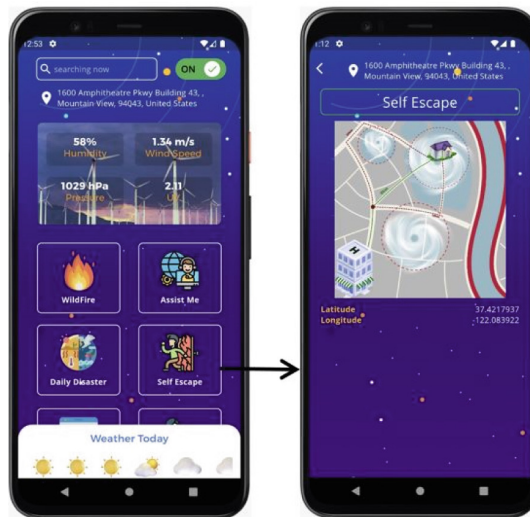
**Fig. 4.** Showing Daily Live Weather.

Figure 3 shows your current precise position, Humidity, wind speed, air pressure, and UV ray index are the four functions shown on the top of the box will display. Humidity refers to the moisture in the air; wind speed refers to the pace at which air moves in your region; atmospheric pressure is a weather indicator; and UV (Ultraviolet) index refers to the intensity of UV radiation. Figure 4 shows the forecast for today based on your location. There are five sorts of weather: sunny, cloudy, windy, rainy, and stormy.

Figure 5 shows the current location of wildfires throughout the world using data from the NASA API & Open weather API. NASA’s satellite sensors are frequently the first to discover wildfires. NASA supplies data that may be used to detect active flames and temperature anomalies, both of which can be seen in real-time. Figure 5 displays your present location as well as information regarding natural disasters that may or may not occur. This application notified an individual about the safe or dangerous zone of this individual’s weather or area. When a person was not in a danger zone, the message “Safe Zone” appeared. If a person’s location in a risk zone is likely to be affected by a natural disaster such as a cyclone, heavy rain or flood, the warning “Danger Zone” appears. Aside from that, there are choices for knowing what is occurring or will happen, as well as safe spots to stay based on your location.



**Fig. 5.** Alerting the persons about the current disaster (Safe Zone and Danger Zone).



**Fig. 6.** Showing a safe route to a person when a nearby disaster happens.

Figure 6 shows the virtual image of a safe location if you are going to discover that a person is in a danger zone. It shows a safe route to a person by tracking his/her real-time location and directs this person to the nearest safe place. Machine learning is a subfield of artificial intelligence (AI) focused on disaster warning systems. Apart from machine learning, some important subfields of AI are soft computing, artificial general intelligence, multi-agent systems, intelligent agents, deep learning, etc. The category of



intelligent agent-based systems is a bit different subfield of AI, such as relative direction learning-based systems [19, 20].

We have only provided some result-based screenshots of the apps in the results section of this study which is not enough for research. Moreover, this study did not discuss the comprehensive data analysis, model validity, data accuracy, or impact of the study which is intended to be discussed in the future study.

## 4 Conclusion

The creation of Wireless Emergency Alerts System, a machine learning system that uses data from the NASA API and OpenWeatherMAP API in both the past and present to forecast natural disasters. To find trends and patterns that might be used to predict future disasters, the study required gathering and analyzing data on previous natural catastrophes, including floods, wildfires, tornadoes, cyclones, and hurricanes. Users of the app may also get GPS-based evacuation routes and details on local shelters in an emergency. By providing access to important resources and information, the app's goal is to assist individuals before, during, and after a natural catastrophe. The software acts as a daily crisis management tool, updating users on the most recent catastrophes and alerting them to them.

Statistical data analysis has become important in this research for accurate disaster forecasting which we plan to include in future studies. Peer-to-peer (P2P) communication is a vital research area for distributed communication. Customer-to-customer (C2C) is a type of P2P that has been used in information quality [16]. Blockchain, on the other hand, is a P2P mechanism that has become necessary in various fields besides cryptocurrency, such as the security of medical information in healthcare [21] and prevention of security attacks [23]. In forthcoming research, blockchain based disaster warning applications may be a focusable area. Finally, this research may become indispensable for disaster management and future enhancement of this research is very important for combating disasters.

**Acknowledgement.** We would like to thank the Bangladesh Association of Software and Information Services (BASIS) for selecting this research as project to compete in the final round of the NASA International Space App Challenge 2022. This honor has motivated us to continue working hard and make a great impact on the world.

## References

1. McBride, S.R., et al.: Latency and geofence testing of wireless emergency alerts intended for the ShakeAlert® earthquake early warning system for the West Coast of the United States of America. *Saf. Sci.* **157**, article no. 105898 (2023)
2. Pillai, A.S., et al.: A service oriented IoT architecture for disaster preparedness and forecasting system. *Internet Things* **14**, article no. 100076 (2021)
3. Snezhana, D.: Applying artificial intelligence (AI) for mitigation climate change consequences of the natural disasters. *Res. J. Ecol. Environ. Sci.* **3**(1), 1–8 (2022)

4. Panahi, M., et al.: Deep learning neural networks for spatially explicit prediction of flash flood probability. *Geosci. Front.* **12**(3), article no. 101076 (2021)
5. Rajeshbabu, S., et al.: Classification of Flood Disasters Severity Levels by Employing Machine Learning Techniques. *EasyChair*, Preprint no. 9570 (2023)
6. Nusrat, F., et al.: A high-resolution earth observations and machine learning-based approach to forecast waterborne disease risk in post-disaster settings. *Climate* **10**(4), article no. 48 (2022)
7. Anderson, L.O., et al.: An alert system for seasonal fire probability forecast for South American protected areas. *Clim. Resilience Sustain.* **1**(1), e19 (2021)
8. Arinta, R.R., WR, E.A.: Natural disaster application on big data and machine learning: a review. In: 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia, pp. 249–254. *IEEE* (2019)
9. Hossain, B., et al.: Climate change induced extreme flood disaster in Bangladesh: Implications on people's livelihoods in the Char Village and their coping mechanisms. *Prog. Disaster Sci.* **6**, article no. 100079 (2020)
10. Kantamaneni, K., et al.: Appraisal of climate change and cyclone trends in Indian coastal states: a systematic approach towards climate action. *Arab. J. Geosci.* **15**, 814 (2022)
11. Xie, M., et al.: Slope disaster monitoring and early warning system based on 3D-MEMS and NB-IoT. In: 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, pp. 90–94. *IEEE* (2019)
12. Ritchie, F., et al.: Natural Disasters. *OurWorldInData.org* (2022). <https://ourworldindata.org/natural-disasters>
13. Sadeq, M.J., et al.: A cloud of things (CoT) approach for monitoring product purchase and price hike. In: Peng, S.L., Son, L.H., Suseendran, G., Balaganesh, D. (eds.) *Intelligent Computing and Innovation on Data Science*. LNNS, vol. 118, pp. 359–368. Springer, Singapore (2020). [https://doi.org/10.1007/978-981-15-3284-9\\_39](https://doi.org/10.1007/978-981-15-3284-9_39)
14. Hasan, M.K., et al.: Evolution of industry and blockchain era: monitoring price hike and corruption using BIoT for smart government and industry 4.0. *IEEE Trans. Industr. Inform.* **18**(12), 9153–9161 (2022)
15. Akhtaruzzaman, M., et al.: HSIC bottleneck based distributed deep learning model for load forecasting in smart grid with a comprehensive survey. *IEEE Access* **8**, 222977–223008 (2020)
16. Haque, R., et al.: Modeling the role of C2C information quality on purchase decision in Facebook. In: Al-Sharhan, S., et al. (eds.) *Challenges and Opportunities in the Digital Era*. I3E 2018. *Lecture Notes in Computer Science*, vol. 11195, pp. 244–254. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-02131-3\\_22](https://doi.org/10.1007/978-3-030-02131-3_22)
17. Sadeq, M.J., et al.: Integration of blockchain and remote database access protocol-based database. In: Yang, X.S., Sherratt, S., Dey, N., Joshi, A. (eds.) *Proceedings of Fifth International Congress on Information and Communication Technology*. *Advances in Intelligent Systems and Computing*, vol. 1184, pp. 533–539. Springer, Singapore (2021). [https://doi.org/10.1007/978-981-15-5859-7\\_53](https://doi.org/10.1007/978-981-15-5859-7_53)
18. Hasan, M.S., et al.: Identification of construction era for Indian subcontinent ancient and heritage buildings by using deep learning. In: Yang, X.S., Sherratt, R.S., Dey, N., Joshi, A. (eds.) *ICICT 2020*. *Advances in Intelligent Systems and Computing*, vol. 1183, pp. 631–640. Springer, Singapore (2021). [https://doi.org/10.1007/978-981-15-5856-6\\_64](https://doi.org/10.1007/978-981-15-5856-6_64)
19. Kabir, S.R., et al.: Relative direction: location path providing method for allied intelligent agent. In: Singh, M., Gupta, P., Tyagi, V., Flusser, J., Ören, T. (eds.) *Advances in Computing and Data Sciences*. *ICACDS 2018*. *Communications in Computer and Information Science*, vol. 905, pp. 381–391. Springer, Singapore (2018). [https://doi.org/10.1007/978-981-13-1810-8\\_38](https://doi.org/10.1007/978-981-13-1810-8_38)

20. Kabir, S.R., et al.: A computational technique for intelligent computers to learn and identify the human's relative directions. In: 2017 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, India, pp. 1037–1040. IEEE (2017)
21. Haque, R., et al.: Blockchain-based information security of electronic medical records (EMR) in a healthcare communication system. In: Peng, S.L., Son, L.H., Suseendran, G., Balaganesh, D. (eds.) Intelligent Computing and Innovation on Data Science. LNNS, vol. 118, pp. 641–650. Springer, Singapore (2020). [https://doi.org/10.1007/978-981-15-3284-9\\_69](https://doi.org/10.1007/978-981-15-3284-9_69)
22. Mitin, S.J., et al.: Identifying COVID-19 pandemic stages using machine learning. In: Chinara, S., Tripathy, A.K., Li, K.C., Sahoo, J.P., Mishra, A.K. (eds.) Advances in Distributed Computing and Machine Learning. LNNS, vol. 660, pp. 231–241. Springer, Singapore (2023). [https://doi.org/10.1007/978-981-99-1203-2\\_20](https://doi.org/10.1007/978-981-99-1203-2_20)
23. Alam, M.R.: Use of blockchain to prevent distributed denial-of-service (DDoS) attack: a systematic literature review. In: Chinara, S., Tripathy, A.K., Li, K.C., Sahoo, J.P., Mishra, A.K. (eds.) Advances in Distributed Computing and Machine Learning. LNNS, vol. 660, pp. 39–47. Springer, Singapore (2023). [https://doi.org/10.1007/978-981-99-1203-2\\_4](https://doi.org/10.1007/978-981-99-1203-2_4)
24. Habib, A.K.M.A., et al.: Distributed denial-of-service attack detection for smart grid wide area measurement system: a hybrid machine learning technique. *Energy Rep.* **10**(9), 638–646 (2023)
25. Ghazal, T.M., et al.: Machine learning approaches for sustainable cities using internet of things. In: Alshurideh, M., Al Kurdi, B.H., Masa'deh, R., Alzoubi, H.M., Salloum, S. (eds.) The Effect of Information Technology on Business and Marketing Intelligence Systems. Studies in Computational Intelligence, vol. 1056. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-12382-5\\_108](https://doi.org/10.1007/978-3-031-12382-5_108)



# Agent-Based Model for Oil Storage Monitor and Control System Using IoT

Hassan Kanj<sup>(✉)</sup>, Abdullah Aljeri, and Tarek Khalifa

College of Engineering and Technology, American University of the Middle East,  
54200 Egaila, Kuwait  
hassan.kanj@aum.edu.kw

**Abstract.** The storage of oil products is characterized by a high level of risk that can be illustrated by an explosion, fire, or spills from the storage area. As a result, monitoring those areas becomes an essential task for stakeholders especially in countries with a wide storage area. This task cannot be monitored manually as this methodology requires a high number of employers and workers. This work proposes a smart IoT-based approach for real-time monitoring and controlling oil storage areas with an evaluation of the risk level. The proposed approach is then modeled and simulated using multi-agent system (ABM) to represent the dynamic behavioral of the oil storage system in normal and degraded mode. Experiments done have demonstrated the effectiveness of ABM in dynamic modeling, the reliability of the proposed system in monitoring and controlling oil storage facilities and emphasizing the advantages of incorporating IoT in oil storage management.

**Keywords:** Agent-based model · IoT · Smart Control System · Oil storage · Monitoring · Risk analysis

## 1 Introduction

Oil has long been the most precious resource that is a part of nearly every aspect of today's world. It can be described as the essence of the global economy, as it is the main energy source for transportation, and industries, as well as serving as a critical raw material for many consumer goods, and it has shaped the geopolitical landscape of our modern world. Oil storage plays a crucial role in the oil and gas sector, as it enables oil companies to safely and efficiently store and transport oil. Besides all safety and security measurements, oil storage presents risk to the surrounding areas, population, and the environment [1]. This risk depends on many factors as type and quantity of products, population density in the storage area, weather conditions and safety barriers implemented in the system. Continuous monitoring of oil storage is essential to guarantee the oil's safety and quality, and to minimize the resulting risk level. The primary objective of oil storage monitoring is to ensure that oil is stored and transported under ideal conditions. This requires keeping track of various factors, such as

temperature, humidity, pressure, and levels of the oil in the tank. By monitoring these parameters, companies can maintain the oil in good condition. Furthermore, monitoring oil storage can help optimize operations and minimize costs. Monitoring of oil conditions and changes allows companies to detect and resolve any emerging issues, like leaks or equipment failures, or a change in parameters before they escalate into significant problems that may generate an environmental damage, air pollution, injuries, and deaths. In California 2015, an explosion of an unmonitored pipeline produces a spill of over 100,000 gallons of oil into the ocean. This incident emphasizes the importance of monitoring oil storage to prevent accidents and the benefits of integrating advanced technologies like IoT for real-time data analysis. Furthermore, such events lead to a change in the behavior of the overall system that can be modeled using Markov model [2], ABM [3], or Petri nets [4]. As ABM can simulate the dynamic behavior of autonomous systems with decentralized entities, it will be used in this work to represent the behavior of the oil storage system with the presence/absence of abnormal readings. The objective of this paper is to develop a smart ABM that constantly monitors and controls oil storage areas and saves all the reading using an IoT platform for analysis and display. The rest of this paper is as follow: Section 2 captures the recent relevant works, Sect. 3 explains the proposed approach. Section 4 represents the implementation done and the obtained results. Finally, Sect. 5 concludes the paper.

## 2 Literature Review

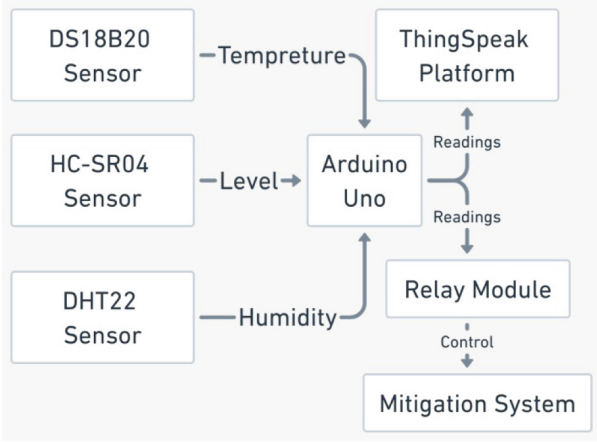
Oil storage units occupy a huge surface especially in US, Canada and in the Gulf region. They may generate serious outcomes to surrounding areas in case of fire, explosion or toxic release. In [5], scientists present an overview of Underground Storage Tank Condition Monitoring Techniques. Authors in [6] developed an intelligent approach for oil storage monitoring that detects and measures any change in the storage area regarding the temperature, gas concentration, and pressure. This approach uses Zigbee technology to communicate all the readings with a controller. If any incident happens, the system notifies the workers via buzzers and text messages. This approach is very cost-effective, but it lacks crucial components like humidity and level sensors, which are vital for monitoring oil parameters. Many works in literature have been done to evaluate and ensure the QoS in mesh network [7], and [8]. Another IOT-based approach proposed in [9], presents a surveillance system using ESP32 cam for fire and smoke detection in oil tanks. The main components used in the system are MQ-2 gas, smoke sensor, IR flame sensor, buzzer, and GSM module, which sends alerts when it detects gas leakage, smoke, or temperature changes. The system also integrates MySQL database [10] for data management and real-time notifications. This system didn't consider the humidity, level, and pressure in the monitoring process. In addition, the system didn't consider any mitigation measures like using a fan to reduce the temperature or the humidity, or a pump to control the level of the tank automatically. Authors in [11] uses Birgé- Massart method to monitor

and reduce the leakage in pipeline, this works lacks of representing the dynamic behavior of the studied system. Computer vision technology has been used in [12] for monitoring the state of oil storage tanks. This technology offers an intelligent analysis of photos high resolution that provide an accurate detection of any defects, cracks in the container but it doesn't show explicitly the direct impact of such events on the performance and safety of the containers. Another approach based AI and machine learning is proposed and reviewed in [13], and [14]. It analyzes the behavior of the main parameters in the storage system to detect and predict future events. In [15] scientists present an oil tank management system that integrates wireless sensor networks (WSN) and the Internet of Things (IoT) to monitor and control oil storage levels, temperatures, and detect fires in real-time. The proposed solution is designed to address the challenges associated with manual monitoring of oil tanks in refineries, those include the risk of fires, floods, and environmental pollution. The proposed system takes advantage of wireless sensor networks and IoT technologies to enable real-time monitoring and control of oil tanks from anywhere. The main drawbacks of the system are the inability of pressure or gas detection, as well as the absence of control components like pump and fans for enhanced safety and efficiency.

### 3 Proposed Model

Monitoring an oil container requires the consideration of all parameters that may affect the quality or generate any undesirable event in the storage area like overload, over-pressure, overheating, corrosion, and erosion [16, 17], and [18]. In the absence of a control system, that makes immediate reaction to mitigate the effects of abnormal readings, the occurrence of such events may produce a critical event that can be represented by a chemical reaction, spills [6], or toxic release.

In this work, an IOT based approach for monitoring and controlling oil storage was proposed. Fig. 1 shows the high-level design of the proposed approach. It uses the following sensors to fully monitor the storage area: DHT-22 air temperature and humidity sensor, a waterproof temperature digital sensor DS18B20 for the oil's temperature, and an ultrasonic HC-SRO4 level sensor. For the control mechanism (mitigation system), a fan and a pump are used to control the temperature and the level in the storage area respectively. This system provides real-time monitoring and control, enhances efficiency and safety by automatically addressing undesired occurrences. All the readings are analyzed and transmitted to ThingSpeak IoT platform through an Arduino UNO microcontroller and an ESP Wi-Fi module. ThingSpeak provides the control center with remote access to the captured data, it can be integrated with a variety of other platforms, including MATLAB [19]. To illustrate the dynamic behavior of the monitored system, a simulation model is needed. Agent-based model (ABM) has been used successfully to analyze risks and represent the dynamic behavior in different applications such as transportation of dangerous good [20], control of agricultural fields [21], monitoring of Elderly People [22].



**Fig. 1.** System's High-level design.

A proposed Agent-based model [23] is used in this work, where every independent unit in the proposed approach is considered as an agent  $A_i$  living in an environment  $\theta_i$  as shown in Fig. 2. Each agent  $A_i$  is described by a functionality, a set of variables ( $V_i$ ), a set of failure modes ( $FM_i$ ) and a set of behavioral modes ( $BM_i$ ) as represented in Eq. 1.

$$A_i = \langle V_i, FM_i, BM_i \rangle \quad (1)$$

When all  $V_i$  are normal,  $FM_i$  will be inactive and consequently the behavioral mode of the agent  $BM_i$  is Nominal. Any change in  $V_i$  may activate  $FM_i$  and as a result  $BM_i$  changes from nominal to degraded.  $BM_i$  remains degraded until  $V_i$  becomes nominal (failure is repaired). This model shows the direct relation between the agent failure mode and behavioral mode [24].

## 4 Implementation and Results

### 4.1 Implementation

The implementation process begins after designing the required circuits, physically connecting all the components, then uploading the code to the Arduino board using the Arduino IDE. And finally testing and checking the overall functionality of the circuits including sensors and actuators to ensure that the designed circuit meets the project requirements. Identifying the type of monitored products is highly important in the experimental process as their properties determine the required conditions for a safe storage regarding the temperature, humidity, and level [25, 26]. The proposed approach was implemented considering a tank of mineral oil with a high flash point located in an area with very low population density. The safe storage conditions of such products are a temperature value (t) in the interval [20 °C, 40 °C], a humidity (h) between 40% to

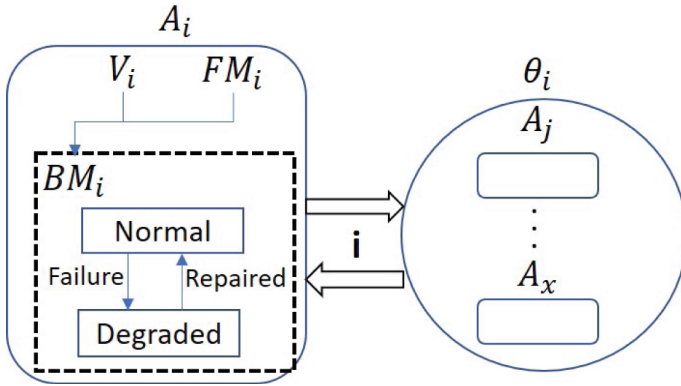


Fig. 2. Multi-agents’ system representation.

60%, and a level (1) in the range 10 to 25 cm [27]. Once the temperature is very close to the maximum threshold (or it exceeds 35 °C), the fan will be activated for 15 s and whenever the oil level drops below 10 cm, the pump refills the tank to 25 cm. Fig. 3 shows the flowchart of the proposed approach and its functionality based on the different scenarios. Fig. 4 illustrates the dynamic behaviour of the system using ABM. Agents’ container and Control center behave normally (in nominal mode) in the absence of abnormal readings (temperature, humidity and level are within the defined ranges). Once any of those parameters becomes abnormal (a failure mode FM is to be activated), and the behavioural mode of the storage area changes to be degraded.

This behavioural mode remains degraded until the failure is repaired (the parameter goes back to the normal range). At the same time, all readings including the information about active failures will be shared with the control center through ThingSpeak. In case of an active failure, the behaviour of the agent “Control center” changes from nominal to degraded until the failure is repaired.

After representing the dynamic behavior of the smart autonomous multi-agents’ system [28], it is very important to analyze the resulting risk level to determine whether it is accepted, or unaccepted. Risk in general is defined as a function of the Likelihood ( $\psi$ ) of an accident and the severity (S) of its outcomes as shown in Eq. 2.

$$RL = F(\psi, S) \tag{2}$$

The likelihood can be studied using a fault tree analysis [29]. The severity depends on the stored products (type and quantity) and the characteristics of surrounding areas (population density, content). After identifying the probability and severity of the accident, the resulting risk level can be represented as shown in Table 1.



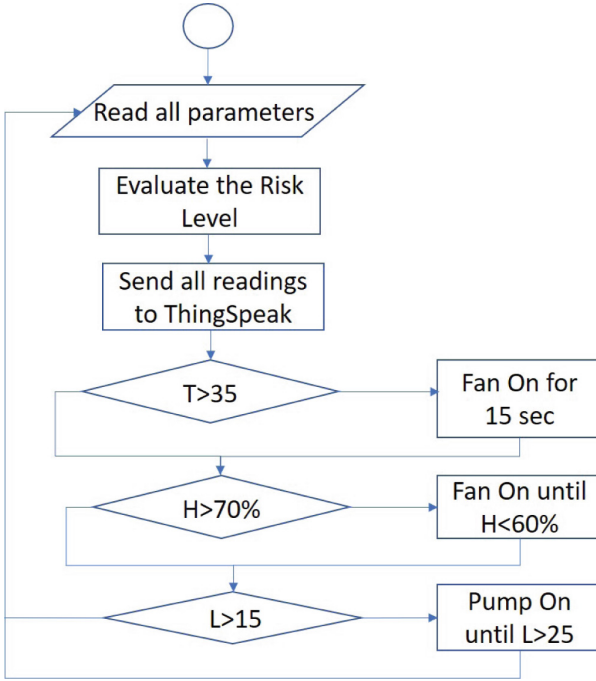


Fig. 3. The system flowchart.

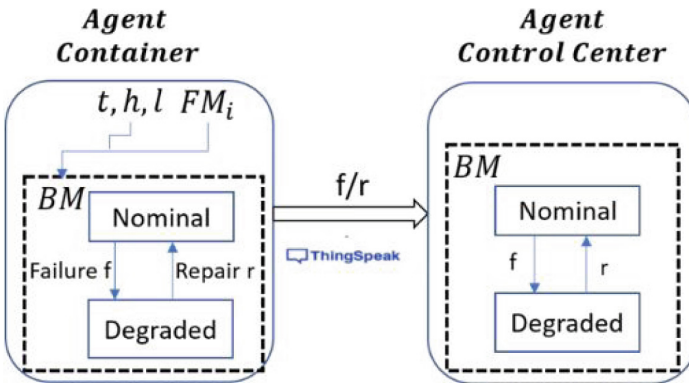


Fig. 4. ABM implementation.

**Table 1.** Risk Level Classification.

Likelihood $\psi$	Severity of Harm				
	Catastrophic	Serious	Medium	Minor	Negligible
Very Likely $\psi \geq 10^{-1}$	high	high	high	significant	significant
Likely $10^{-3} \leq \psi < 10^{-1}$	high	high	significant	significant	moderate
Moderate $10^{-6} \leq \psi < 10^{-3}$	high	high	significant	moderate	low
Unlikely $10^{-9} \leq \psi < 10^{-6}$	high	significant	moderate	low	low
Rare $\psi < 10^{-9}$	significant	significant	moderate	low	low

## 4.2 Results

Many experiments were done to check the functionality of the proposed approach in different conditions (with and without abnormal reading/failures). First sensors were tested individually then experiments were applied for the whole system. Table 2 presents 8 consecutive readings provided by the component DS18B20. The first four were done in a nominal environment (22 °C for room temperature oil) then the second four in a degraded one (45 °C for hot oil).

**Table 2.** Sensors’ Readings in nominal and degraded modes.

Liquid temperature (C)	Air temperature (C)	Humidity (%)
22.00	21.10	52.90
22.00	21.10	51.90
22.00	21.10	51.90
22.00	21.10	50.50
45.00	26.30	92.60
45.00	26.30	92.60
45.00	26.30	92.90
45.00	26.30	92.90

Next, a couple of tests were done to test the efficiency of the HC-SRO4 sensors as shown in Table 2. The first test involved placing the sensor in an oil container nearly full, whereas the second test was done with the same container being almost empty. Results show that the sensor was able to measure the liquid levels effectively with an output of 8 cm and 3 cm for the almost empty and the nearly full container respectively. Next, readings of the environment temperature and humidity were collected in two environments: inside and outside apartment.

Outputs were as follows: 21 °C and 49% humidity, and later 27 °C and 90% humidity in the hot container.

All the readings were immediately shared with the control center through ThinkSpeak. ThinkSpeak interface is used to represent the variation and the current values of the monitored parameters. Table 3 represents the dynamic behavior of the agents in the system in addition to the status of the actuators reaction. If all the monitored parameters are normal, both agent' behavioral modes are Nominal and both actuators are Off. In case of abnormal reading, agents' behavioral mode change to degraded and the needed action is triggered (fun/pump on) until the reparation of the failure. Experiments done under many conditions to show an important change in all parameters and the effect of such change on the Agents' behavioral mode.

**Table 3.** Dynamic Behavior of the System' agents with evaluation of the risk level.

Temperature (t°) Humidity (h), Level (l)	Agents' behavior			Risk Level
<b>t = 21.1, h = 52.90%, l = 13</b>	Nominal	Nominal	Pump: Off, Fan: Off	moderate
<b>t = 21.1, h = 52.90%, l = 3.8</b>	Degraded	Degraded	Pump: On, Fan: Off	significant
<b>t = 43, h = 48.80%, l = 9</b>	Degraded	Degraded	Pump: On, Fan: On	significant
<b>t = 41, h = 65%, l = 14.4</b>	Degraded	Degraded	Pump: Off, Fan: On	significant
<b>t = 37, h = 50.0%, l = 14</b>	Nominal	Nominal	Pump: Off, Fan: Off	moderate

## 5 Conclusion

This paper presents a smart Agent-based model for Oil Storage Monitor and Control system using IoT, where the objective is to represent the dynamic behavior of an oil storage system in nominal and degraded modes and to automatically control any undesirable change in the storage environment. Two agents were defined: agent container and agent control center with many behavioral and failure modes that are directly related to the sensors' readings. If all the readings are normal, agents' s behavioral modes are nominal. With the presence of abnormal readings, the agent' behavioral mode is set to degraded. The connection between the agents was established using ThinkSpeak IOT platform. Experiments prove that the use of accurate and efficient components improves the real-time detection of any abnormal condition. Furthermore, ABM provides a full representation of the dynamic behavior of the agents in the storage system. It shows the direct relation between the abnormal readings, agent' failure mode and the agent' behavioral mode. It describes how an active failure may affect the behavior of the agents in the system. As future work, the proposed model can be extended to evaluate risks of oil storage in addition to the representation of the failure propagation in the agent model.

## References

1. Kanj, H., Flaus, J.-M.: An agent-based framework for mitigating hazardous materials transport risk. In: 2015 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS), pp. 1–8. IEEE (2015)
2. Brameret, P.-A., Rauzy, A., Roussel, J.-M.: Automated generation of partial Markov chain from high level descriptions. *Reliab. Eng. Syst. Saf.* **139**, 179–187 (2015)
3. Kanj, H., Aly, W.H.F., Kanj, S.: A novel dynamic approach for risk analysis and simulation using multi-agents model. *Appl. Sci.* **12**(10), 5062 (2022)
4. Taleb-Berrouane, M., Khan, F., Amyotte, P.: Bayesian stochastic petri nets (BSPN)-a new modelling tool for dynamic safety and reliability analysis. *Reliab. Eng. Syst. Saf.* **193**, 106587 (2020)
5. Sheng, O.C., Ngui, W.K., Hoou, H.K., Hee, L.M., Leong, M.S.: Review of underground storage tank condition monitoring techniques. In: MATEC Web of Conferences, vol. 255, pp. 02009. EDP Sciences (2019)
6. Prasad, B., Manjunatha, R.: Internet of things based monitoring system for oil tanks. In: 2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC), pp. 1–7. IEEE (2021)
7. Ashtaiwi, A., Hassanein, H.: Utilizing IEEE 802.11n to enhance QoS support in wireless mesh networks. In: 2008 33rd IEEE Conference on Local Computer Networks (LCN), pp. 689–696 (2008)
8. Ashtaiwi, A., Saoud, A., Almerhag, I., et al.: Performance evaluation of VANETs routing protocols. *Comput. Sci. Inf. Technol.* **4**, 305–315 (2014)
9. Rajasekaran, A., Chetan, T., Naveen, T., Sai, V.B.: IoT based smart oil and gas monitoring system (2020)
10. Li, P., Zhang, N., Yang, L.: Oil test liquid level monitoring system based on the IoT. In: 2022 4th International Conference on Intelligent Control, Measurement and Signal Processing (ICMSP), pp. 429–433. IEEE (2022)
11. Gao, J., Ai, B., Hao, B., Guo, B., Zheng, Y.: A boosted wavelet improvement thresholding algorithm based on birgé-massart strategy for pipeline leakage signal noise reduction processing. In: 2022 4th International Conference on Intelligent Control, Measurement and Signal Processing (ICMSP), pp. 738–742 (2022)
12. Sharapov, A., Bugakova, T.Y., Basargin, A.: Application of computer vision technology for monitoring the condition of oil storage tanks. In: *Journal of Physics: Conference Series*, vol. 2032, no. 1, p. 012097. IOP Publishing (2021)
13. Ashtaiwi, A.: Artificial intelligence is transforming the world development indicators. In: 2020 IEEE / ITU International Conference on Artificial Intelligence for Good (AI4G), pp. 122–128 (2020)
14. Huby, A.A., Sagban, R., Alubady, R.: Oil spill detection based on machine learning and deep learning: a review. In: 2022 5th International Conference on Engineering Technology and its Applications (IICETA), pp. 85–90 (2022)
15. Chandla Ellis, R., Dharshini, D., Yashwanthi, M.: An IoT based system for improving the solar water heater. *Math. Stat. Eng. Appl.* **71**(4), 5840–5853 (2022)
16. Fingas, M.: Introduction to oil chemistry and properties. *Handbook of Oil Spill Science and Technology*, pp. 51–77 (2014)
17. Aalsalem, M.Y., Khan, W.Z., Gharibi, W., Armi, N.: An intelligent oil and gas well monitoring system based on internet of things. In: 2017 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET), pp. 124–127. IEEE (2017)

18. Al-Humairi, S.N.S., Ravindran, D.R.T., Abdullah, M.I., Hamzah, B., Alkawaz, M.H.: Intelligent monitoring system for oil well efficiency. In: 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA), vol. 2020, pp. 13–17. IEEE (2020)
19. Li, S., Zhang, C., Xu, D., Pan, Y., Ni, J., Qiao, Y.: A novel on-line system for monitoring the oil level of capsule-type storage tank of power transformer. In: 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2), pp. 1–9. IEEE (2018)
20. Kanj, H., Kotb, Y., Flaus, J.-M.: A proposed petri-net extension to analyze risk for evolving systems using an agent model. In: 2020 International Conference on Control, Automation and Diagnosis (ICCAD), pp. 1–7 (2020)
21. Humoud, T., et al.: Smart agriculture monitoring and controlling system using multi-agents model. In: 2023 5th International Conference on Bio-engineering for Smart Technologies (BioSMART), pp. 1–4 (2023)
22. Aldousari, A., et al.: A wearable IoT- based healthcare monitoring system for elderly people. In: 2023 5th International Conference on Bio-engineering for Smart Technologies (BioSMART), pp. 1–4 (2023)
23. Kanj, H.: Contribution to risk analysis related to the transport of hazardous materials by agent-based simulation, Ph.D. dissertation, Université Grenoble Alpes (2016)
24. Yang, D., Zhang, X., Hu, Z., Yang, Y., Oil contamination monitoring based on dielectric constant measurement. In: International Conference on Measuring Technology and Mechatronics Automation, vol. 1, pp. 249–252. IEEE (2009)
25. Kanj, H., Flaus, J.-M.: A meta model framework for risk analysis, diagnosis and simulation. In: Safety and Reliability Conference ESREL (2014)
26. Abbod, A.A., Zwayer, N.B.: Using internet of things techniques to measure parameters of oil tanks. *J. Pet. Res. Stud.* **11**(1), 153–167 (2021)
27. Pavithra, D., Balakrishnan, R.: IoT based monitoring and control system for home automation. In: Global Conference on Communication Technologies (GCCT), pp. 169–173. IEEE (2015)
28. Kotb, Y., Alakkoummi, M., Kanj, H.: Reinforcement learning based framework for real time fault tolerance. In: 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 0357–0364. IEEE (2020)
29. Kabir, S.: An overview of fault tree analysis and its application in model based dependability analysis. *Expert Syst. Appl.* **77**, 114–135 (2017)



# A Novel Method to Detect High Impedance Fault in Electric Vehicle Integrated Distribution System

Pampa Sinha<sup>1</sup>, S. Ramana Kumar Joga<sup>2</sup>, Kaushik Paul<sup>3</sup>,  
and Fausto Pedro García Márquez<sup>4</sup>(✉)

<sup>1</sup> School of Electrical Engineering, KIIT Deemed to be University, Bhubaneswar, India  
pampa.sinhafel@kiit.ac.in

<sup>2</sup> Department of EEE, Dadi Institute of Engineering and Technology, Anakapalle, India

<sup>3</sup> Department of Electrical Engineering, BIT Sindri, Dhanbad 828123, India

<sup>4</sup> Ingenium Research Group, University of Castilla-La Mancha, Ciudad Real, Spain  
faustopedro.garcia@uclm.es

**Abstract.** A fault is a typical state that alters the functioning of the power grid. An abrupt alteration in the current level is noted in the event of a fault due to the incorporation of the Electric Vehicle Charging Station. The distribution system is subject to various types of faults, including open circuit faults and short circuit faults such as L-L-L-G, L-G, L-L-L, and L-L-G faults. The majority of traditional over-current relays are capable of detecting these categories of faults. The over-current relays exhibited a deficiency in detecting high-impedance faults, as their fault current was found to be lower than the standard current value. The presence of high impedance faults can result in the occurrence of arcs and the potential for fire arcing within the distribution system. This results in significant loss of both property and human life. Accurately and promptly detecting these types of faults is imperative. This paper employs a new approach utilizing RADWT & SVM to identify and categorize high-impedance faults in the electrical distribution network. The utilization of the rational dilation wavelet transform is employed for the purpose of feature extraction, while the support vector machine is utilized for the detection and classification of faults.

**Keywords:** First Keyword · Second Keyword · Third Keyword

## 1 Introduction

The prevalence of electrical distribution systems contributes to an elevated probability of faults occurring within them. According to reference [1], an electrical fault refers to an atypical condition that can arise from equipment failure within the power system, unfavorable weather patterns, or human error. The potential consequences of such errors may range from harm to individuals to equipment impairment, contingent on the gravity of the situation. In the event of a fault, there is a deviation from the nominal voltage and current values, leading to an overabundance of either high or low current flow within the

distribution system. This leads to the damage of machinery and equipment, in addition to the loss of human lives. Prompt initiation of fault identification is necessary to minimize the duration of the blackout. Within an electrical distribution system, various types of faults may arise, such as L-G, L-L, L-L-G, and high-impedance faults. The timely detection of malfunctions is imperative for ensuring the reliability of the power grid and mitigating the potential hazards to human life and property. Over-current relays of conventional type encounter challenges in detecting faults in high-impedance loads such as this. High impedance faults (HIF) occur when an electrical conductor establishes electrical connectivity with a surface such as a road, a tree limb, or an object that impedes the flow of fault current to a level that is not reliably detected by conventional over-current protection devices. This may result in Hypoxia-Inducible Factor (HIF) activation. Currently, professionals are directing their efforts towards devising techniques for detecting high-impedance faults as a means of enhancing the reliability of the power distribution infrastructure [2]. The wavelet-based feature extraction approach is frequently utilized in practical applications for analyzing transient signals. In this study, the Rational Dilation Wavelet Transform (RADWT) is employed for feature extraction and transient signal analysis. Several scholars suggest specific algorithms for identifying and categorizing high-impedance faults in distribution systems. The high-impedance fault data sets were formatted in a standard hierarchical data format by Douglas Gomes, as documented in reference [3]. The detection and classification of high-impedance faults in distribution systems that are integrated with photovoltaic systems are achieved through the utilization of a novel signal processing technique referred to as Hermit transform, as expounded upon in reference [4]. Reference [5] provides a comprehensive overview of high-impedance faults. The article [6] presents a method for diagnosing high-impedance faults in an electrical distribution system using an empirical wavelet transform and differential faulty energy-based approach. In [7], Katleho Molo presented a method for detecting high-impedance faults through the utilization of DWT and SVM-based techniques. The proposed method was verified using a decision tree algorithm, and its findings were discussed. The article [8] presents a discussion on a distribution network's high-impedance fault method that relies on a semi-supervised detection algorithm. In reference [9], Pallav Kumar Bera has put forth a method for detecting high-impedance faults in microgrids through the utilisation of machine learning. The article referenced as [10] presents a discussion on the utilisation of a differential scheme that is sensitive to time domain for the purpose of detecting and categorizing high-impedance faults. In reference [11], Hua Ding presented a distinct algorithm for identifying high-impedance faults within a multi-terminal HVDC system. The article referenced as [12] discusses the inadequacy of current techniques that rely on travelling wave and energy distribution for detecting and categorizing high-impedance faults.

The present research work employs the rational dilation wavelet transform for the purpose of signal decomposition. The utilisation of a support vector machine classifier has been employed for the purpose of detecting and classifying high impedance faults (HIF) from low impedance faults (LIF) in the distribution system. The primary contribution of this study is to mitigate the limitations associated with the discrete wavelet transform with respect to phase localization. The proposed method is also evaluated with consideration given to noise signals.

## 1.1 Modelling of High Impedance Fault

Extensive research efforts have been dedicated to the modelling of high-impedance faults, leading to a consensus among scholars that such faults exhibit nonlinearity and asymmetry. Emanuel formulated an HIF model, which integrates empirical observations with theoretical constructs. It is recommended to utilize a configuration consisting of two direct current sources that are coupled antiparallel with two diodes in order to attain zero periods of arcing and asymmetry. The presence of uneven resistances in the system represents the flow of asymmetric fault currents. The symbols  $R_P$  and  $R_N$  are utilized to represent the resistances of faults. The provision of non-equal values to the two resistances facilitates the production of asymmetric fault currents. The Emmanuel method has been subject to further modifications through the incorporation of a sequence of anti-parallel diodes, variable resistance, and control voltage sources. Figure 1 depicts the block diagram of the HIF arc model that has been put forward.

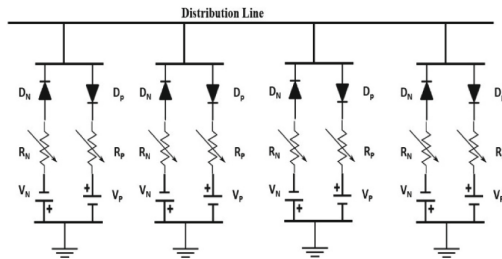


Fig. 1. Proposed HIF Model

## 2 Methodology

Feature extraction is a mathematical technique employed to reduce high-dimensional input datasets into a lower dimension without sacrificing any of the essential information contained in the original dataset. This can be classified as a mathematical methodology. The process of feature extraction is a crucial element of the protective system employed to improve fault categorization. Feature extraction methods serve as a fundamental tool for addressing the classification and regression challenges prevalent in power systems. These methodologies are efficacious instruments. Feature extraction techniques are employed to reduce the dimensionality of a complex data set while retaining its essential properties. This enables the data to be more easily analyzed and processed. The present study retrieved two statistical parameters from the recovered signal, namely the energy level and the entropy level of the signal. A classification is a quantitative instrument utilized to identify specific attributes of information from a vast array of features. The task is achieved through the process of analyzing and highlighting the similarities and differences of the characteristics. The utilisation of this phenomenon has been widely employed in the domain of power system condition monitoring. The development of protection systems with precise pattern recognition capabilities is imperative as it facilitates



prompt execution of protection responses in the occurrence of fault scenarios. Following the fault event, the procedure utilizes the measured initial values of the fault current at the terminal of the source node. Subsequently, the RADWT technique is employed to transmit the ascertained actual current information through the utilisation of low-pass and high-pass filters. Figure 2 displays the visual representation of the proposed algorithm in the form of a block diagram.

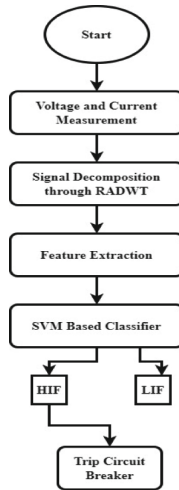


Fig. 2. Flowchart of Proposed HIF detection Algorithm

### 2.1 Decomposition of Voltage and Current Signals Through Rational Dilation Wavelet Transform

The methodology employed for the dissection and amalgamation of acoustic oscillations for the purpose of defect analysis, which relies on the redundant discrete wavelet transform (RADWT), bears resemblance to the approach that rests on the dyadic wavelet transform. In contrast, the rational-dilation wavelet transform adopts a sampling strategy that exhibits high density in both the time and frequency domains. Empirical evidence suggests that the processing or sifting of oscillatory signals within a restricted time frame, such as fault signals, necessitates a superior frequency resolution than that which is provided by the dyadic wavelet transform. RADWT has been developed by utilising iterated two-channel filter banks in conjunction with the dilation factor. This approach enables RADWT to possess both the shift-invariant property and the ability to handle discrete-time data. The mathematical expression yields a rational integer which can be utilized to denote the dilation factor of the redundant discrete wavelet transform (RADWT). [13]

$$Dilation\ Factor = \frac{a}{b} \tag{1}$$

$$Redundancy = \frac{1}{c} \left[ \frac{1}{1} - \frac{a}{b} \right] \tag{2}$$

$$Q - factor = \sqrt{ab} \left[ \frac{1}{1} - \left( \frac{a}{b} \right) \right] \quad (3)$$

## 2.2 Support Vector Machine Algorithm for Fault Detection

The objective of utilizing support vector machines (SVMs), also referred to as SVM classifiers, is to identify a hyperplane that serves as a margin of separation between two distinct classes of data. The process involves the projection of input vectors onto a high-dimensional space, followed by the computation of the hyperplane. Typically, a hyperplane is deemed to be in an optimal state when it fulfils two conditions. Two key requirements for effective classification are: ensuring a clear demarcation between data classes through a wider margin and maximizing the distance between the hyperplane and the nearest data class [14]. Thus, it is feasible to compute the optimal hyperplane through the determination of a resolution to the quadratic programming problem presented by

$$\text{minimum} \frac{1}{2} \|w\|^2 + D \left( \sum_{l=1}^L \rho_l \right) \quad (4)$$

$$\text{Subjected to } m_i(w \cdot x_i + b) \geq 1 - \rho_i$$

The variables  $w$  and  $b$  is given as

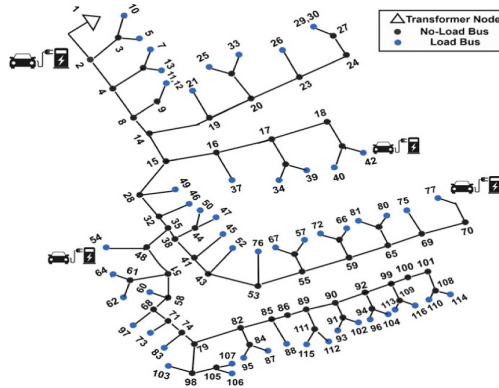
$$w = \sum_{l=1}^n \zeta_l y_l x_l \quad (5)$$

$$b = y_{\text{support vectors}} - \sum_{i=1}^l \zeta_i y_i m(x, x_i) \quad (6)$$

## 3 Results and Discussions

The proposed method was evaluated using the IEEE European Low Voltage Test Feeder. The present test cases are focused on North American standard systems; however, low-voltage distribution systems, encompassing both radial and meshed configurations, are extensively prevalent in regions beyond North America. Ensuring that the tools encompass both prevalent modes of distribution system configuration is imperative. The objective of this test case is to address a gap in the current benchmarks by providing a range of low-voltage configurations that are commonly encountered. In the realm of distribution research and planning, there is a growing demand for time-series solutions that can effectively capture the dynamic behaviour of various technologies over the medium- to long-term, as opposed to static power flow solutions. This represents a departure from the preceding scenario, wherein it was posited that static power flow resolutions would prove satisfactory. In order to achieve a precise evaluation of concepts or commodities such as Volt Var Optimisation, coordinated regulator and capacitor controls, energy storage, or

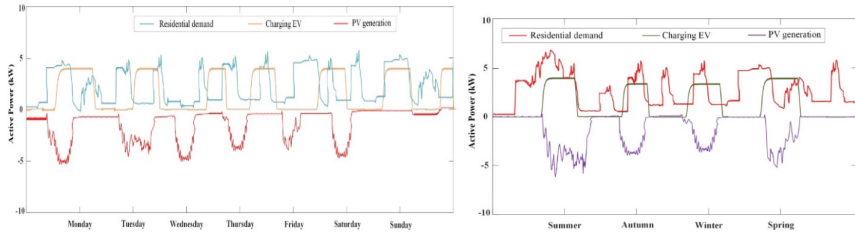
solar, a certain duration of time must elapse prior to attaining a complete understanding of their functionality. In order to meet the necessary criteria, a test case for low voltage in Europe was formulated, possessing the subsequent attributes: The test feeder exhibits a voltage level of 416 V (phase-to-phase), which is a prevalent characteristic of low voltage distribution systems commonly observed in Europe. The time-series simulation utilises load forms that provide a temporal resolution of one minute throughout a twenty-four-hour period. This paper presents the outcomes of a time-series simulation conducted over a 24-h period, along with the findings of a static power flow analysis performed at various significant intervals. The radial distribution feeder of the low-voltage test feeder exhibits a fundamental frequency of 50 Hz. The feeder and the medium voltage (MV) system are interconnected through a transformer located at the substation. The voltage conversion process involves the utilisation of a transformer to transform the high voltage of 11 kV into a lower and safer voltage of 416 V. The voltage of all power lines, including the primary feeder and any branch lines, is established at 416 V. Figure 3 depicts the Line pictorial representation of the IEEE European Low Voltage Test Feeder that has been proposed.



**Fig. 3.** Single Line Diagram of the proposed IEEE European Low Voltage test feeder integrated with EV charging stations

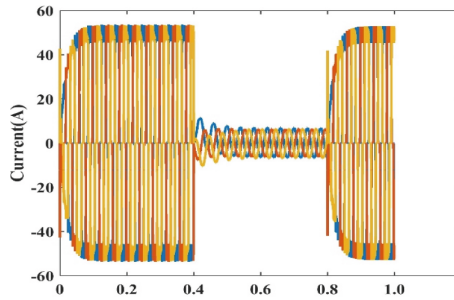
The EV Charging station serves as the proximal load, while the PV generation functions as the proximal source. The integration of Electric Vehicles into the distribution grid has been implemented at specific nodes, namely node 1, node 54, node 42, and node 77. Figure 4 displays the simulated plot of the weekly load demand. Figure 4 illustrates that the demand for residential and electric vehicle charging on Monday is approximately equivalent to the power generated by photovoltaic generation. It has been observed that EV charging stations serve as an active load to the distribution system on a daily basis throughout the week. On Sunday, the photovoltaic generation exhibited a significantly reduced output due to the low demand from residential and electric vehicle charging station loads. The load demand analysis has been conducted for the entire year, and it has been graphically represented in Fig. 4. The seasonal variation of photovoltaic (PV) generation has been noted, with higher levels occurring during the summer months and

lower levels during the winter months. The load on both residential and electric vehicle charging stations tends to be higher during the summer season and lower during the winter season. It has been observed that the electric vehicle (EV) charging station constitutes an active load consistently throughout the year. Figure 3 depicts the introduction of a high-impedance fault at node 77.



**Fig. 4.** Weekly and Yearly load demand

The arc current at node 77 is shown in Fig. 5.



**Fig. 5.** Arc Current at node 77

Figure 5 illustrates that the fault current exhibits asymmetry and is lower than the rated load current. The presence of non-linearity has been detected in the waveform of the current. It has been observed that fault current exhibits stochastic variations over time. The fault current undergoes decomposition via the Rational Dilation Wavelet Transform, as depicted in Fig. 6.

Figure 6 illustrates the decomposition of HIF Fault Current through RADWT at 19 levels. In this context, the Quality factor has been determined to be 6, while the values of P and S are 5 and 2, respectively. The redundancy value has been computed to be approximately 3.0, while the dilation factor has been estimated to be around 1.20. An elevated quality factor enhances the likelihood of detecting a subtle deviation from the baseline signal. Given that the dilation factor exceeds 1, it has been observed that the image resolution frequency is satisfactory. The present study involves the utilisation of decomposed fault current signals to extract the necessary features that are subsequently fed into a supervised machine learning classifier based on Support Vector Machines.

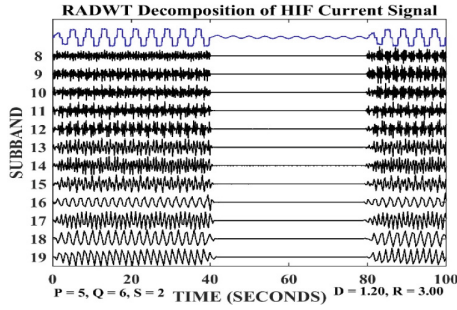


Fig. 6. RADWT Decomposition of HIF Current Signal

The purpose of this approach is to facilitate the classification process. The fault current signals were subjected to decomposition into frequency components, with a maximum of 19 levels of decomposition. The training dataset comprises 2000 frequency components. Figure 7 displays the scatter plot of fault classification.

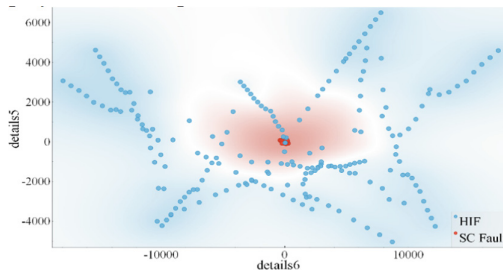


Fig. 7. Scatter Diagram of Fault Classification

A power system operating in real-time is subject to varying degrees of noise. The performance of the proposed method is also impacted by the presence of noise. The method under consideration was evaluated in the presence of noise levels of 40 dB and 25 dB signal-to-noise ratio (SNR). A noise condition is considered to be present in the fundamental fault signal when the SNR is at a level of 25 dB. A signal-to-noise ratio (SNR) of 40 dB can be considered a low noise or negligible noise scenario in relation to the fundamental fault signal. The mathematical expression for the noise condition formula is as follows:

$$SNR_{dB} = 20 \ln_{10} \left( \frac{I_{HIF}}{I_{noise}} \right) \tag{7}$$

where  $I_{HIF}$  is HIF Current Signal and  $I_{Noise}$  is Noise Signal. The Performance of HIF Detection and Classification for various Kernel SVM at no noise and noise conditions is tabulated in Table 1.

**Table 1.** High impedance fault classification performance analysis

CLASSIFIER	KERNEL	PQD's CLASSIFICATION ACCURACY (%)	
		SNR = 40 dB Normal	SNR = 25 dB Noisy
SVM	Linear	75.00	62.50
	Gaussian	99.50	95.00
	Polynomial	99.00	93.75
	Sigmoid	98.50	92.50
	RBF	99.75	96.25

The RBF SVM classifier demonstrated a classification accuracy of approximately 99.75% and 96.25% in the presence of noise. The proposed method is also evaluated with the methods that exist in the literature and it is tabulated in Table 2.

**Table 2.** Overall comparison of fault classification

Method	Accuracy (%)	Condition	Dependability (%)
Proposed Method [ RADWT + RBF SVM]	99.75	Noise and Normal	100
DWT + SVM	99.34	Normal	-
DWT + DT	98.22	Normal	95.79
Stockwell + ANN	99.15	Normal and Noise	-
Morphology Gradient + Fuzzy Logic Classifier	99.40	Normal	99.78
DFT + ANFIS	99.64	Normal	-

## 4 Conclusions and Limitations

In this paper, the efficiency in detecting high impedance faults is better with the RADWT-based signal processing method than with the traditional Fourier Transform-based wavelet transform. In contrast to a normal wavelet transform, the Quality factor can be made bigger or smaller. Because of this, the suggested method was better at finding and labelling High-impedance Faults that happen in distribution systems that include charging points for electric vehicles. Fault detection is done with an SVM classifier that has different kernel functions. The results show that the RBF kernel-based SVM classifier is the best way to find and sort high-impedance problems. The proposed method is tried on real-time IEEE European Low Voltage distribution under both normal and busy situations. We can say that for 25 dB SNR, the proposed way worked better. RADWT

needs more levels of decomposition, which makes the proposed method a little slow. This is thought to be the proposed method's biggest flaw. In terms of classification accuracy and reliability, the proposed method also did better than other methods in the books.

## References

1. Joga, S.R.K., Sinha, P., Maharana, M.K., Jena, C.: Tunable Q-wavelet transform based entropy measurement to detect and classify faults in certs microgrid test bed. In: 2021 International Conference in Advances in Power, Signal, and Information Technology (APSIT), pp. 1–6 (2021). <https://doi.org/10.1109/APSIT52773.2021.9641113>
2. Joga, S.R.K., Sinha, P., Maharana, M.K.: A novel graph search and machine learning method to detect and locate high impedance fault zone in distribution system. *Eng. Rep.* **5**(1), e12556 (2023). <https://doi.org/10.1002/eng2.12556>
3. Gomes, D.P.S., Ozansoy, C.: VeHIF: an accessible vegetation high-impedance fault data set format. *IEEE Trans. Power Delivery* **37**(6), 5473–5475 (2022). <https://doi.org/10.1109/TPWRD.2022.3195002>
4. Guillen, D., Olveres, J., Torres-García, V., Escalante-Ramírez, B.: Hermite transform based algorithm for detection and classification of high-impedance faults. *IEEE Access* **10**, 79962–79973 (2022). <https://doi.org/10.1109/ACCESS.2022.3194525>
5. Ghaderi, A., Ginn, H.L., III., Mohammadpour, H.A.: High impedance fault detection: a review. *Electr. Power Syst. Res.* **143**, 376–388 (2017)
6. Gao, J., Wang, X., Wang, X., Yang, A., Yuan, H., Wei, X.: A high-impedance fault detection method for distribution systems based on empirical wavelet transform and differential faulty energy. *IEEE Trans. Smart Grid* **13**(2), 900–912 (2022). <https://doi.org/10.1109/TSG.2021.3129315>
7. Moloi, K., Davidson, I.: High-impedance fault detection protection scheme for power distribution systems. *Mathematics* **10**, 4298 (2022). <https://doi.org/10.3390/math10224298>
8. Guo, Z.-Y., Guo, M.-F., Gao, J.-H.: High-impedance fault semi-supervised detection of distribution networks based on tri-training and support vector machine. In: 2022 IEEE 3rd China International Youth Conference on Electrical Engineering (CIYCEE), pp. 1–6 (2022). <https://doi.org/10.1109/CIYCEE55749.2022.9958950>
9. Bera, P.K., Kumar, V., Pani, S.R., Bargate, V.: Detection of high-impedance faults in microgrids using machine learning. In: 2022 IEEE Green Energy and Smart System Systems (IGESSC), pp. 1–5 (2022). <https://doi.org/10.1109/IGESSC55810.2022.9955330>
10. Chisava, O., Ramos, G., Celeita, D.: Time-domain sensitive differential protection approach for high-impedance faults (HIF). *IEEE Ind. Appl. Soc. Annual Meeting (IAS)* **2022**, 1–6 (2022). <https://doi.org/10.1109/IAS54023.2022.9940107>
11. Ding, H., Huai, Q., Wang, J., Zhu, S., Li, Y., Hu, Z.: A novel high-impedance fault recognition method for multi-terminal HVDC system. In: 2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT), pp. 1–6 (2022). <https://doi.org/10.1109/GlobConPT57482.2022.9938185>
12. Feng, D., Hongfei, S., Xiangjun, Z., Fan, X., Hang, Z., Yihan, Z.: High-impedance fault detection method based on sparsity of traveling wave full waveform energy distribution. In: China International Conference on Electricity Distribution (CICED), pp. 1566–1570 (2022). <https://doi.org/10.1109/CICED56215.2022.9929011>

13. Jog, S.R.K., Sinha, P., Maharana, M.K.: Artificial intelligence in classifying high-impedance faults in electrical power distribution system. In: Proceedings of International Conference on Recent Trends in Computing, Communication & Networking Technologies (ICRTCCNT) 2019 (2019). <https://ssrn.com/abstract=3430316>, <https://doi.org/10.2139/ssrn.3430316>
14. Hemamalini, S.: Rational-dilation wavelet transform based torque estimation from acoustic signals for fault diagnosis in a three-phase induction motor. *IEEE Trans. Ind. Inform.* **15**(6), 3492–3501 (2018). <https://doi.org/10.1109/TII.2018.2874463>





# Multi-class Classification of Voice Disorders Using Deep Transfer Learning

Mehtab Ur Rahman<sup>(✉)</sup>  and Cem Direkoglu 

Electrical and Electronics Engineering, Middle East Technical University - Northern Cyprus Campus, North Cyprus, via Mersin 10, Kalkanli, Guzelyurt 99738, Turkey  
{mehtab.rahman,cemdir}@metu.edu.tr

**Abstract.** Voice disorders are a widespread issue affecting people of all ages, and accurate diagnosis is crucial for effective treatment. With the recent development of artificial intelligence-based audio and speech processing, research on detection and classification of voice disorders has increased. However, existing work has mostly focused on the binary (two class) classification of voice disorders. Some researchers have also explored multi-class classification, but their results are not promising. In this paper, a framework is proposed for the multi-class classification of voice disorders using OpenL3 embeddings. A pre-trained OpenL3 model is utilized to extract high-level embedding features from the mel spectrogram. Then different classifiers are evaluated after the neighbourhood component analysis (NCA) based feature selection. Random Forest (RF), Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) are employed separately to classify the selected features. The evaluation and comparison are performed on a balanced subset of the Saarbruecken voice database (SVD). Without any speech enhancement preprocessing, our best model, OpenL3-KNN improves the existing work accuracy by 4.9% and F1 score by 8.7%.

**Keywords:** Voice disorder · Multi-class classification · OpenL3

## 1 Introduction

The acoustic characteristics of speech voices are typically analyzed by loudness, quality and pitch [17]. The term “Voice disorders” refers to a medical condition that affects the quality, pitch, and loudness of an individual’s voice, often caused by a dysfunction in the laryngeal, respiratory system, or vocal tract. These disorders can reduce the clarity of a person’s oral communication ability. Voice disorders can vary in severity from slight hoarseness to complete voice loss [1]. Voice disorders are commonly classified based on their etiology, including psychogenic, functional or organic in nature. Organic voice disorders are characterized by structural or neurological abnormalities that affect the functioning of voice production organs [16]. Functional voice disorders arise from the inefficient utilization of the vocal mechanism despite the normal physical structure of the

larynx and vocal tract. Psychological conditions such as sadness, anxiety, and emotional reactions to traumatic or stressful situations can all be the root cause of psychogenic voice disorders [2].

Voice disorders may cause a significant impact on people of all ages, which can lead to stress, embarrassment, frustration, withdrawal, and depression. People in certain professions or activities, such as teachers, telephone salespersons, lawyers, tour guides, actors and singers are particularly vulnerable to voice disorders [15]. To provide appropriate treatment, it is important to accurately classify the type of voice disorder. A speech therapist evaluates the patient's voice quality for the identification of voice disorders. However, this approach of evaluation is subjective and is dependent on the expertise of the speech therapist. Another approach to evaluate voice disorders is the Artificial Intelligence (AI) based processing of voice signal acoustic features, which provides an objective assessment [3]. Particularly with the development of deep learning techniques, significant advancement observed in the AI based decision making systems. It is important to note that the AI based objective evaluation approach should only be used for initial assessment, and the final decision regarding the diagnosis of voice disorder should be made by medical professionals with expertise in the field. Researchers' interest in automating this procedure has grown in recent years, which not only offers a faster identification tool for speech therapists but also makes it more comfortable for patients.

The rest of the paper is organized as follows: Sect. 2 reviews related work that is based on computational intelligence. Section 3 describes the dataset used in our experiments. Section 4 explains the proposed work. Section 5 presents our experiments and results. Finally, Sect. 6 concludes the paper.

## 2 Related Work

The evaluation of pathological voices has been the focus of various studies in the literature. Specifically, these studies concentrate on extracting specific features from acoustic voices and evaluating the ability to classify between different voice disorders based on these features. A multi-level classification approach was used in [4] to assess voice disorders using four binary classifiers trained with support vector machines (SVM). The study explored glottal parameters, perturbation measures, OpenSMILE and Mel-Frequency Cepstral Coefficients (MFCC) as source features, and evaluated the approach using the Saarbruecken Voice Database (SVD). In [5], ensemble learning was proposed using two independent classifiers, Gaussian Mixture Model (GMM) and SVM with features obtained from the MFCC and Modulation Spectra (MS) respectively. The authors in [6] developed an algorithm for the identification of laryngeal pathologies. In [7], the authors proposed a feature space transformation technique for pathological voices detection. They used Hidden Markov Models to transform the feature space of Mel-Frequency Cepstral Coefficients and short-term noise parameter features. Authors in [8] extracted features from voice signals using the Hilbert-Huang Transform and then used KNN as a classifier. Their accuracy on the VOICE ICar fEDerico II (VOICED II) database was 93.3%.

In recent years, the rapid advancement of deep learning has led to excellent performance in various fields. As a result, many researchers have attempted to utilize it to the classification of voice disorders. In [9], MFCC features were extracted from voice signals and utilized as input to a deep neural network (DNN). In [10], the authors introduced a novel algorithm for the detection of pathological voices. They used a Convolutional Neural Network (CNN) for feature extraction from spectrograms of audio signals. To make the system more robust, they also used a Convolutional Deep Belief Network to initialize the weights. In [12], a system was proposed called CS-PVC for voice disorders detection and classification. The system first extracts MFSC features from the voice signals before using DCA-ResNet to predict voice disorders. [13] analyzed an algorithm that extracts a chromagram acoustic feature from voice samples and uses it as input for a CNN-based classification system. The suggested approach achieved 85% accuracy for classification.

In [20], the authors presented a pre-trained OpenL3-SVM transfer learning framework utilizing linear local tangent space alignment (LLTSA) for dimension reduction. The framework utilizes Mel spectrum features and demonstrates promising results in the classification of voice disorders. In paper [14], the authors employ a Deep Neural Network (DNN), as well as Support Vector Machine (SVM) and Random Forest (RF) classifiers, along with three handcrafted features - signal energy (SE), zero-crossing rates (ZCRs), and signal entropy (SH) - for the multi-label classification of voice disorders. In [11], the authors presented a method that utilizes pre-trained self-supervised models (wav2vec 2.0 and HuBERT) as feature extractors and employs a hierarchical classifier for the multi-label classification of voice disorders. In [18], the authors focused on modeling and fine-tuning a deep learning architecture of CNN and ResNet34 layers for voice disorders detection. The proposed system achieved an accuracy of 95.41%. [19,21,22] used VGG-16 for voice disorders detection.

It is noteworthy that most of the studies have concentrated on the binary classification of voice disorders i.e. the classification between pathological voices and healthy voices. Others have focused on identifying specific pathological voices among all other pathological and healthy voices. A few researchers have investigated multi-class classification of voice disorders; however, the accuracy of such approaches is significantly lower. The current state of the art techniques for multi-class classification of voice disorders require further improvement to enhance their accuracy. To bridge this research gap, we employed a pre-trained OpenL3 model with various classifiers and Neighborhood Component Analysis (NCA) based feature selection, aiming to enhance the accuracy of multi-class classification for voice disorders.

In this paper, we propose a novel three-stage framework for the multi-class classification of voice disorders. Firstly, the voice signals are sliced into chunks and then frequency and time masking is applied to allow data augmentation. In the second stage, high-level embedding features are extracted from mel spectrograms using the pre-trained OpenL3 model [23]. After neighbourhood component

analysis (NCA) based feature selection, various classifiers are employed in the third stage. Additionally, we conducted a comparative analysis of our framework with existing methods to assess its performance.

### 3 Dataset

For this research, we selected a subset of voice samples from the publicly available “Saarbruecken Voice Database” (SVD) [24]. A total of 2,041 voice recording sessions from 681 healthy and 1,019 pathological people are included in the database. Individuals who were recorded before their pathology and recorded after recovery were included in both the healthy and pathological categories. Sessions are recorded at 50 kHz sampling frequency and a resolution of 16 bits. For each recording, the voice signal and the EGG signal have been. The voice recording session consists of the German phrase “Guten Morgen, wie geht es Ihnen?” which means “Good morning, how are you?” and the vowels ‘a’, ‘i’, and ‘u’ pronounced at four different pitches (low, normal, high and rising-falling). In our experimental tests, we selected four pathologies: Chronic laryngitis, Vocal fold polyp, Vocal cord paresis and Functional dysphonia. For each pathology, we included 38 recordings, 19 of which were from male speakers and 19 from female speakers. The recordings for these four pathologies have been selected with the vowel ‘a’ in normal pitch. We aimed to ensure a well-balanced dataset. The reason for selecting these specific pathologies was to compare our experimental results with existing studies. Table 1 provides details of the dataset used in this research.

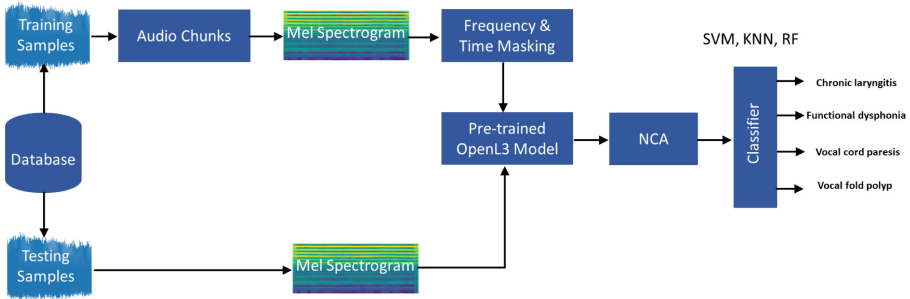
**Table 1.** Quantity of voice recordings for each pathology class.

Pathology	Male Recordings	Female Recordings	Total Recordings
Chronic laryngitis	19	19	38
Functional dysphonia	19	19	38
Vocal cord paresis	19	19	38
Vocal fold polyp	19	19	38

### 4 Proposed Method

Figure 1 provides an illustration of the proposed classification system for voice disorders in this study. In our study, we implemented several data preprocessing steps and feature selection. Each audio signal was divided into 0.5s segments, audio segments that were less than 0.3s were excluded. For chunks between 0.3 and 0.5s of duration, zero padding was applied to ensure that all audio segments have the same length. To extract meaningful features from the audio signals, we calculated the mel spectrogram. In order to calculate the mel spectrogram for

each audio segment, the Short-Time Fourier Transform (STFT) is computed with a Hamming window of 30 ms and a shift of 15 ms. The design of the mel spectrogram is based on psychoacoustic analysis, aimed at approximating the human auditory system. The spectrogram is converted to decibel units using the logarithmic scale.

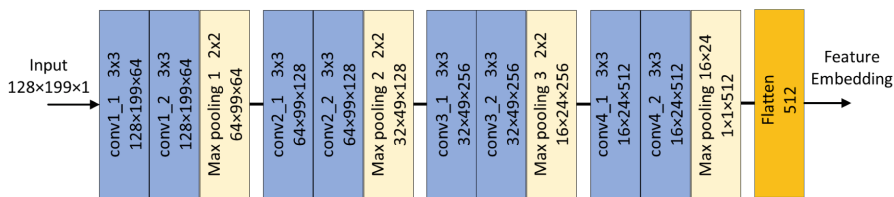


**Fig. 1.** The Proposed Voice Disorders Classification System

We used frequency masking and time masking data augmentation to improve the performance of our voice disorder classification system. The frequency masking technique randomly masks a range of consecutive frequency bands in the mel spectrogram to introduce variability and enhance the model's ability to generalize to new data. The time masking technique randomly masks a portion of consecutive time steps in the mel spectrogram to introduce temporal variation and improve the model's robustness to small temporal shifts in the data.

We used the pre-trained OpenL3 model to extract high-level embedding features. OpenL3 [23] is the extension of  $L^3$ -Net. OpenL3 is a deep learning-based audio embedding model that can be used to extract rich deep audio embeddings of audio signals. For training the embedding models, OpenL3 uses a music subset and an environmental subset of AudioSet. Human, animal and other sounds heard in a natural acoustic environment are included in the environmental subset. The model was trained using a supervised learning approach to map audio signals to high-dimensional embeddings. OpenL3 computes HTK Mel-spectrograms with 128 and 256 Mel bands as input to the embedding models. The output embeddings can be either 6144 or 512 dimensional.

The architecture of OpenL3 is shown in Fig. 2. Batch normalization is applied to the input layer and each convolution layer, and ReLU activation function is applied to every batch normalization layer except for the first one. To extract feature embeddings, the classification layer of the OpenL3 network is removed. OpenL3 offers pre-trained models that can be used to extract embeddings for various types of audio signals, including music, environmental sounds, and speech. We extracted 512 -dimensional high-level embedding features with pre-trained OpenL3 models with the music subset and mel256 as a Spectrogram representation.



**Fig. 2.** OpenL3 Structure

Neighborhood Components Analysis (NCA) [25] is used to reduce the dimensionality of the OpenL3 embeddings. NCA is a supervised learning algorithm that is used for dimensionality reduction, which can help improve classification performance and reduce training time. The original embedding vector has a high dimensionality of 512. We applied NCA to transform it into a new 25-dimensional, at captures the most important patterns and relationships in the data, which helped us to improve the model performance.

We trained and tested three machine learning classifiers: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF). SVM is a popular classification algorithm that is widely used in machine learning. SVM is based on statistical learning theory. It is highly effective in handling data with high dimensions, as it can find a hyperplane in the dimensional space that effectively separates the samples into different classes [26]. KNN algorithm, also known as a lazy learner, is a straightforward machine learning algorithm. It uses a distance measure to identify the K-nearest neighbors in the training data for a given instance. The final decision is made by analyzing the class labels of the K-nearest neighbors [27]. RF classifier is an ensemble classifier that uses multiple decision trees, each generated using a randomly selected subset of training examples and attributes. The combination of these trees allows for robust predictions, as each tree contributes a unit vote towards the majority class to classify a test instance [28].

## 5 Experiments and Results

In this study, a 70/30 data split was used for training and testing, respectively, ensuring no data leakage occurred from training to testing. All three classifiers used the same training set and test set in their training and testing process. The system performance was evaluated based on accuracy and F1 score. Accuracy is the ratio of the number of correctly classified instances to the total number of instances in the dataset Eq. (1). Here, the variables TP and TN represent the number of True Positives and True Negatives, respectively, whereas FP and FN represent the number of False Positives and False Negatives, respectively. The F1 score is computed as the harmonic mean of precision and recall Eq. (2). To obtain a reliable estimate of the performance, the classifiers were tested 10 times, and the average accuracy and F1 score with standard deviation were recorded.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (2)$$

SVM classifier was trained with a polynomial kernel function of degree 3, regularization parameter 1 and 'scale' kernel coefficient. In KNN classifier, we use K=7 nearest neighbors, and the algorithm used to compute the nearest neighbors is "auto" with uniform weights. Euclidean distance is used as the distance metric. RF classifier was trained with 300 decision trees.

Depending on the subset of databases used, the classification performance of voice disorders can vary from study to study. In our study, we compared the performance of our framework to a previous study [3], which also utilized the same subset from the SVD database. Our proposed OpenL3-KNN method outperforms OpenL3-SVM, OpenL3-RF, and other existing techniques. It should be noted that in [3], speech enhancement techniques were also utilized to enhance the voice data. Table 2 illustrates our system's performance and compares it with [3]. Table 3 shows the precision, recall, and F1 score for each pathology with OpenL3-KNN.

**Table 2.** Performance Comparison. Boldface Values Indicate the Best Performance.

Method	Accuracy± std	F1 Score± std
CNN-LSTM [3]	49.26 ± 2.32	44.3 ± 0.018
OpenL3-SVM (Proposed)	43.75 ± 0.0212	44.17 ± 0.0241
OpenL3-RF (Proposed)	43.75 ± 0.0441	43.13 ± 0.0431
OpenL3-KNN (Proposed)	<b>54.16±0.0276</b>	<b>53.02 ± 0.0345</b>

**Table 3.** Precision, Recall and F1 Score for Each Pathology with OpenL3-KNN.

Pathology	Precision	Recall	F1 Score
Chronic laryngitis	42.11	66.67	51.61
Functional dysphonia	57.14	66.67	61.54
Vocal cord paresis	70.00	58.33	63.64
Vocal fold polyp	60.00	25.00	35.29

Although our proposed method demonstrates superior performance compared to other methods, it is important to acknowledge certain limitations that should be addressed for the practical application of this research. There is a need to enhance the accuracy of the method to ensure its robustness in real-world scenarios. Additionally, it would be valuable to test the method on other pathological voices to assess its generalizability.

## 6 Conclusions

We have presented a three-stage multi-class classification system for voice disorders. In the first stage, voice signals are sliced into chunks and then frequency and time masking are applied. A pre-trained OpenL3 model is used in the second stage to extract high-level embedding features from the mel spectrogram. Finally, SVM, RF and KNN are experimented as classifiers after NCA-based feature selection. A balanced subset of SVD of four types of voice disorders with the vowel “a” in normal pitch was used for training and testing. Results demonstrate that OpenL3-KNN outperforms OpenL3-SVM and OpenL3-RF, and the other existing technique. However, voice disorder classification is still a challenging task mainly because of limited data available for training and testing. In this study, we did not employ any speech enhancement technique as a pre-processing stage. Pre-processing techniques could be explored in future research to improve classification performance. Additionally, further experiments on different datasets could be conducted to evaluate the generalization ability of the proposed system.

## References

1. Ramig, L.O., Verdolini, K.: Treatment efficacy: voice disorders. *J. Speech Lang. Hear. Res.* **41**(1), S101–S116 (1998)
2. American Speech-Language-Hearing Association. Voice disorders.[Practice Portal]. Accessed 30 Dec 2021. <https://www.asha.org/Practice-Portal/Clinical-Topics/Voice-Disorders>
3. Chaiani, M., Selouani, S.A., Boudraa, M., Yakoub, M.S.: Voice disorder classification using speech enhancement and deep learning models. *Biocybern. Biomed. Eng.* **42**(2), 463–480 (2022)
4. Barche, P., Gurugubelli, K., Vuppala, A.K.: Towards automatic assessment of voice disorders: a clinical approach. In: *INTERSPEECH*, pp. 2537–2541 (2020)
5. Arias-Londoño, J.D., Godino-Llorente, J.I., Markaki, M., Stylianou, Y.: On combining information from modulation spectra and mel-frequency cepstral coefficients for automatic detection of pathological voices. *Logoped. Phoniatr. Vocol.* **36**(2), 60–69 (2011)
6. Fonseca, E.S., Guido, R.C., Scalassara, P.R., Maciel, C.D., Pereira, J.C.: Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders. *Comput. Biol. Med.* **37**(4), 571–578 (2007)
7. Arias-Londoño, J.D., Godino-Llorente, J.I., Sáenz-Lechón, N., Osmá-Ruiz, V., Castellanos-Domínguez, G.: An improved method for voice pathology detection by means of a HMM-based feature space transformation. *Pattern Recogn.* **43**(9), 3100–3112 (2010)
8. Chen, L., Wang, C., Chen, J., Xiang, Z., Hu, X.: Voice disorder identification by using Hilbert-Huang transform (HHT) and K nearest neighbor (KNN). *J. Voice* **35**(6), 932–e1 (2021)
9. Chen, L., Chen, J.: Deep neural network for automatic classification of pathological voice signals. *J. Voice* **36**(2), 288–e15 (2022)
10. Wu, H., Soraghan, J., Lowit, A., Di Caterina, G.: A deep learning method for pathological voice detection using convolutional deep belief networks. In: *Inter-speech 2018* (2018)



11. Tirronen, S., Kadiri, S.R., Alku, P.: Hierarchical multi-class classification of voice disorders using self-supervised models and glottal features. *IEEE Open J. Signal Process.* **4**, 80–88 (2023)
12. Ding, H., Gu, Z., Dai, P., Zhou, Z., Wang, L., Wu, X.: Deep connected attention (DCA) ResNet for robust voice pathology detection and classification. *Biomed. Signal Process. Control* **70**, 102973 (2021)
13. Islam, R., Tarique, M.: A novel convolutional neural network based dysphonic voice detection algorithm using chromagram. *Int. J. Electr. Comput. Eng.* (2088–8708) **12**(5) (2022)
14. Junior, S.B., Guido, R.C., Aguiar, G.J., Santana, E.J., Junior, M.L.P., Patil, H.A.: Multiple voice disorders in the same individual: investigating handcrafted features, multi-label classification algorithms, and base-learners. *Speech Commun.* 102952 (2023)
15. Ribas, D., Pastor, M.A., Miguel, A., Martínez, D., Ortega, A., Lleida, E.: Automatic voice disorder detection using self-supervised representations. *IEEE Access* **11**, 14915–14927 (2023)
16. Robotti, C., et al.: Treatment of relapsing functional and organic dysphonia: a narrative literature review. *Acta Otorhinolaryngol. Ital.* **43**(2 Suppl 1), S84 (2023)
17. Schenck, A., Hilger, A.I., Levant, S., Kim, J.H., Lester-Smith, R.A., Larson, C.: The effect of pitch and loudness auditory feedback perturbations on vocal quality during sustained phonation. *J. Voice* **37**(1), 37–47 (2023)
18. Mohammed, M.A., et al.: Voice pathology detection and classification using convolutional neural network model. *Appl. Sci.* **10**(11), 3723 (2020)
19. Vavrek, L., Hires, M., Kumar, D., Drotar, P.: Deep convolutional neural network for detection of pathological speech. In 2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMII), pp. 000245–000250. IEEE (2021)
20. Peng, X., Xu, H., Liu, J., Wang, J., He, C.: Voice disorder classification using convolutional neural network based on deep transfer learning. *Sci. Rep.* **13**(1), 7264 (2023)
21. Gumelar, A.B., Yuniarno, E.M., Anggraeni, W., Sugiarto, I., Mahindara, V.R., Purnomo, M.H.: Enhancing detection of pathological voice disorder based on deep VGG-16 CNN. In: 2020 3rd International Conference on Biomedical Engineering (IBIOMED), pp. 28–33. IEEE (2020)
22. Zakaria, S., Thanush, S., Mugilan, M.: Voice disorder identification using convolutional neural network. In 2022 1st International Conference on Computational Science and Technology (ICCST), pp. 923–927. IEEE (2022)
23. Cramer, J., Wu, H.H., Salamon, J., Bello, J.P.: Look, listen, and learn more: design choices for deep audio embeddings. In: ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3852–3856. IEEE (2019)
24. Woldert-Jokisz, B.: Saarbruecken voice database (2007). [https://www.stimmdatenbank.coli.uni-saarland.de/help\\_en.php4](https://www.stimmdatenbank.coli.uni-saarland.de/help_en.php4)
25. Goldberger, J., Hinton, G.E., Roweis, S., Salakhutdinov, R.R.: Neighbourhood components analysis. In: *Advances in Neural Information Processing Systems*, vol. 17 (2004)
26. Vapnik, V.: *The Nature of Statistical Learning Theory*. Springer, Cham (1999)
27. Alpaydin, E.: *Introduction to Machine Learning*. MIT press, Cambridge (2020)
28. Belgiu, M., Drăgu, L.: Random forest in remote sensing: a review of applications and future directions. *ISPRS J. Photogramm. Remote. Sens.* **114**, 24–31 (2016)



# Coronary Artery Blockage Detection by Automated Segmentation of Vessels in X-Ray Angiograms

Jayanthi Ganapathy<sup>1</sup> (✉), Fausto Pedro García Márquez<sup>2</sup>, and C. H. Dhamini<sup>1</sup>

<sup>1</sup> Faculty of Engineering and Technology, Sri Ramachandra Institute of Higher Education and Research, Chennai, Tamil Nadu, India

{jayanthig, e0120031}@sret.edu.in

<sup>2</sup> Ingenium Research Group, University of Castilla-La Mancha, Ciudad Real, Spain  
Faustopedro.garcia@uclm.es

**Abstract.** Coronary heart disease leading up to stenosis, the partial or total blocking of coronary arteries, is the leading cause of death worldwide. Multi-vessel coronary artery disease affecting two or more coronary requires interpretive expertise on the assessment, the process of interpreting is complex and a time-consuming. Auto-mated identification and classification of angiograms with blockage detection from minimally invasive procedures would be of great clinical value. This study aims at OpenCV method with the help of adaptive thresholding, and brightness corrections for the segmentation of the blood vessels and used the same masked images for the detection of blockage in x-ray angiograms using deep neural nets. The proposed model have achieved 97% accuracy on the task of classifying the x-ray angiogram to 'Blockage Detected' and 'Normal', with a F1-Score of 0.9532. These results open the way to a fully automated method for the identification of Blockage from X-ray angiograms.

**Keywords:** Machine Learning · Coronary Artery · Deep Neural Network · X-Ray Angiogram · Healthcare

## 1 Introduction

Coronary artery disease a non-communicable disease which is affecting a million of people. The main cause of coronary heart disease is atherosclerotic plaque accumulation in the epicardial arteries leading myocardial oxygen supply and demand and resulting in serve pain is the most likely common symptom that occurs during stress either physical or emotional.

Invasive coronary angiography is the preferred tool to assess the complex coronary artery disease, where by a dye is inserted into the patients and a sequence of X-ray images are taken, as the dye is flushed into the coronary arteries as the dye passes through the blood vessels making the blood flow clearer for taking the clear images and then which we can use them for the blockage detection and stenosis detection. It's typically hard to approximate stenosis severity via visual inspection. So, cardiologists

use revascularization procedures such as stent placement, for stenosis detection, using a diameter of blood vessel if the area is less than its diameter of the blood vessel i.e., if the reduction area is greater than 70% then it's believed to be stenosis, which leads to serving angiogram, which can be served by using angioplasty according to the condition of the patient and the type of blood vessel where the blockage is created.

Deep learning is a subset of Artificial Intelligence and is a branch of Machine Learning. It is developed by taking inspiration from the functions and the structure of the human brain. Algorithms are created and the main use of them is understand and interpret huge amounts of complex data, which is not so efficient with the help of Machine Learning. These algorithms (artificial neural networks), are built up of layers of nodes which are interconnected just like the neurons in our brains, which are trained with large data amounts.

Deep learning can also handle real-world problems i.e. when it comes to the performance of algorithms in solving complex operations, Deep learning algorithms are better. In terms of feature extraction, we would have to manually feed the features to Machine Learning algorithms but this is not required when it comes to Deep Learning algorithms as it automatically generates the features related to objects of interest. Deep learning has a wide application range in medical fields detection of stenosis, and tumour cells, other applications include self-driving cars, translations, robotics and many more.

The most popular and special type of network is used in image identification and classification and image clustering. These are designed to act on images or objects with grid structures in them. Each layer in this network consists of convolutional filters. These filters are applied to incoming data to extract patterns and features like shape, size, edges and corners. It can have multiple layers and each of them is interconnected and learn how to detect different characteristics of the input. Since they can automatically generate features from input images, they are widely used in image detection and classification.

## 2 Related Work

Analysis of angiograms is not an easy task using X-Ray images, the difficulties that are faced in the processing and analysis of the images are, Shape of the blood vessels, Density and diameter of blood vessels for the detection of stenosis and the blockage, The background noise in the image is relatively high, so it's difficult to differentiate between the tissues, blood vessels and makes hard for segmentation of the blood vessels for the analysis.

Although many methods have been used for blood vessel segmentation, which can be classified as pattern recognition, model-based tracking, and propagation according to the method [1–3] they have used for segmentation. Most of the blood vessel detection was done using Gaussian matched filter [4]. An adaptive thresholding method using Gaussian filters [5] to extract vessels from X-Ray angiogram. Model-based approaches for medical image segmentation were also applied [6]. Several filter-based approaches [8–11] have been developed for blood vessel segmentation, but the limitation of these works is that they are unable to trap/find the vessel when the bone or a light is overlapped on the region, to overcome this few have tried on convolutional neural Networks (CNN) for segmentation. Many models were built on CNN for segmentation [12–15], and few

studies have used U-Net [16], U-Net is a type of CNN model for biomedical segmentation, with a simple architecture and produces high accuracy, the main advantage of this model is that it can be trained on small datasets of hundreds of images due its simple architecture and makes it well suited for medical imaging applications. [12] Developed a Deep neural nets model using U-Net for segmentation of major branches of the coronary arteries but not all for all the branches.

In [7] author proposed a Deep neural net model named AngioNet for the detection of the blood vessels, where they created a pipeline model for image pre-processing and segmentation, it's an automatic angiographic vessel segmentation for coronary stenosis in the clinical workflow. They have built their own Deep Nets architecture to overcome all the above-mentioned limitations, which combines an Angiographic Processing Network (APN) with a semantic segmentation network. This model contains three main stages Angiographic Processing Network, Deeplabv3 +, U-Net or fully convolutional networks (FCNs). As mentioned above APN is used for semantic segmentation, to address challenges faced during angiographic segmentation, like low contrast images, overlapping bones and the distribution of the light, Deeplabv3 + (39, 40) as the backbone for the segmentation and U-Net as network backbone for comparison. But they haven't given the proper dataset for validation and they haven't mentioned how the trained masks were derived for training they just did segmentation of the angiograms, but they didn't detect whether there is a blockage present or not, if two images were passed to the model and one image is just a normal heart image.

To overcome all the shortcomings, we have built a model, such that if an angiogram X-Ray image with a blockage is passed to the model it predicts where the possible blockage is present which was built using OpenCV using few thresholding methods and applying contours for the image. Although the predictions or correct but it finds the possible blockage. We also built a Deep nets model for the detection of whether the passed X-Ray image has any blockages present or not, furthermore this work will be continued for stenosis detection and the percentage of the blockage which can be achieved by using few pre-trained models like Faster R-CNN, U-Net, ResNet-54, ResNet-150 and many more and any one of the models can be taken based on the performance. As of now CNN model was developed, MobileNet models have been used for the detection whether the passed image have any Blockage or it's just a normal image.

### 3 Methodology

In this project, the first step was to acquire the required dataset. The dataset that was obtained in the form of Video, which was pre-processed and frames were obtained and for the obtained frames pre-processing is done.

After the images were obtained, we used OpenCV for the detection of possible blockages, so first Image Corrections were done like augmentation, Tilting, Orientation, Normalization and corrections of brightness. After that we applied different types of thresholding methods, like adaptive thresholding, Otsu, and Watershed after that we applied contours for the obtained image for thresholding. Using contours, we were able to detect the possible blockages present in the given X-Ray angiogram, although it didn't give the optimal results for the image.

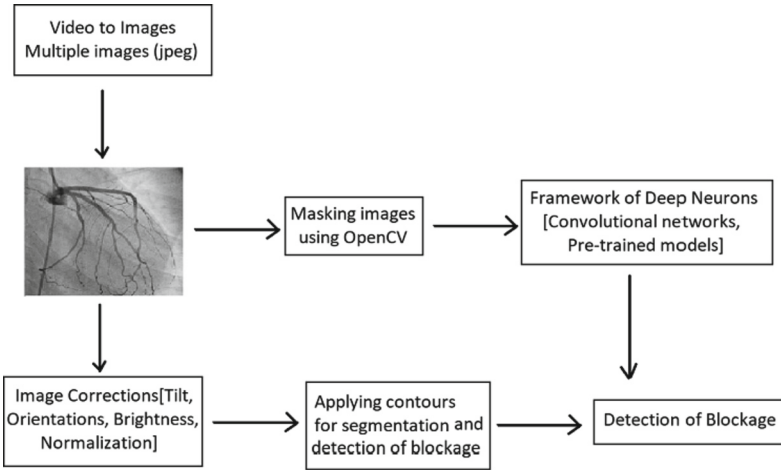


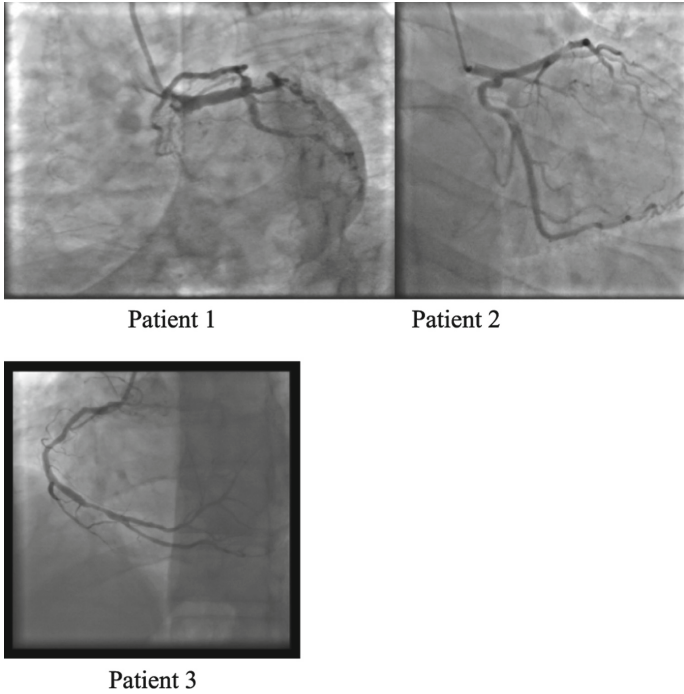
Fig.1. Workflow Diagram

Deep Neural Nets like Convolutional Neural Network (CNN) which was defined by our own for obtaining the optimal results. We even used a few pre-trained models like MobileNet, ResNet-50, and ResNet-154, but only a few of them gave an optimal result. For training the model a masked image but not the X-Ray image the main idea behind just passing the masked is to avoid unnecessary data fed to the model, through which the run time increases. The masked images were obtained using few OpneCV techniques as mentioned above, and converted the X-Ray image to binary image, and removed the noise and corrected brightness by equalizing the brightness throughout the image and applied Gaussian filters for obtaining the blood vessels. Although all the images were not segmented properly because of the backbone overlapping and spotlight. Deep Neural Nets were used but unable to find any datasets which contains both original and masked images for training the dataset. Hence, OpenCV is used for masking. Once masking is done images are passed to the model for the prediction of the image whether the image contains the blockage or it's a normal image. We developed a dashboard for diagnosing the tool we saved the model using trained weights and loaded the model for fast output in the dashboard.

A web based dashboard was developed using Streamlite App. When an angiogram X-Ray is uploaded it gives the output whether the given image have any blockage or not. Figure 1 shows the workflow of the project.

#### 4 Experimental Design: Results and Discussion

The data was curated by a doctor from SRMC (Sri Ramachandra Medical Centre) who is an expert cardiologist, and the data was obtained from patients' treated in the same hospital. The data consists of ten patients' video angiograms from all the possible directions, in that few of the patients have blockage and few of the patients don't have the



**Fig. 2.** Sample X-Ray images from the data

blockage present. The video is obtained in the form of .mvi format. The data was properly anonymized to preserve participant privacy. All participants were over the age of thirty-five.

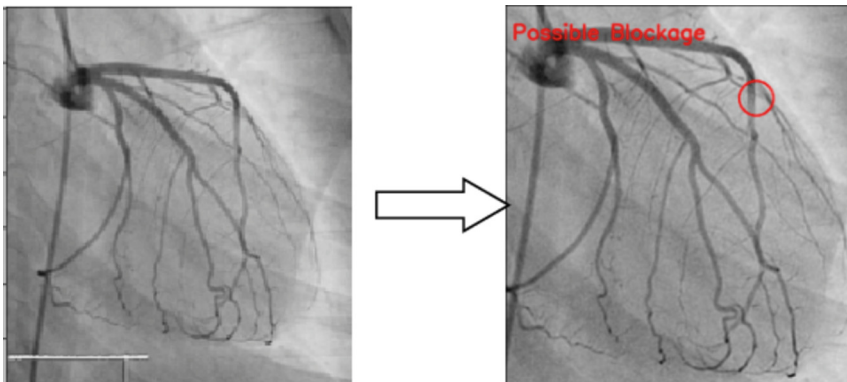
Angiographic images of the radiopaque overlaid coronary arteries with stenotic segments were selected. The selected set of videos was pre-processed, i.e., a given video is converted set of frames using Python. We got around 40 angiogram videos, out of them 10 videos are not used for stenosis detection, they are normal heart images in which no blockage is detected. After converting all the videos to frame around 1234 images were obtained of  $512 \times 512$  pixels grey-scale images were included. Of them 510 (50%) images were used for training the data and the remaining images were stored for further study. We have even used the dataset from Mendeley for stenosis detection, since the doctor hasn't given any annotations for the data. But so far, we have used only the hospital dataset which is collected from SRMC for the detection of possible blockages in the past images. Figure 2 shows you the obtained data from angiogram videos of the size  $512 \times 512 \times 1$ . Further the selected data is divided in the ratio of 80:10:10, i.e., 70% (408 images) is used for training the model, 10% for validation data, and 10% as testing data.

Machine learning algorithms were applied to detect whether there is a possible blockage present. Totally, five different models with different architecture and number of weights. Before getting started with machine learning algorithms we used OpenCV for the possible blockage detections.

To obtain the above results we first did few pre-processing steps like removing the borders, the images are trimmed by 10% from all the sides to remove the noise from the image and converting the image to gray scale, since its already in the form of  $512 \times 512$ , the new shape of the image will be  $512 \times 512 \times 1$ , where the third parameter represents the color channel. To adjust the brightness of an image, changed the value of all pixels by a constant, and added the positive pixels where they are already in dark color, so the segmentation of blood vessels will be easier and subtracted a positive constant from all the pixel values to make the image darker.

The Median Blur operation is similar to the averaging methods. Here every element has replaced the median of all the surrounding pixels. Although there are many other spatial filters, median blur value doesn't create a new unrealistic pixels when the filter straddles an edge. Using median filter we have removed the noise from the images. It's a very crucial in the image processing field as it removes the noise as well as preserves sharp edges simultaneously.

Thresholding is a type of image segmentation, where we change the pixels of an image to make the image easier to analyze, there are many techniques involved in thresholding such as Simple Adaptive Thresholding, Adaptive Thresholding, Otsu, Gaussian adaptive thresholding, Watershed and so on. Adaptive thresholding gave a clear output for analyzing the image, when used the block size as 3. Finally, contours are applied after thresholding. Contours joins all the points along the boundary of an image having the same intensity, through which a masked image is obtained, which will be easy for analyzing the stenosis. After that we applied dilation and erosion for the images, and compared the images. If the pixels matches, then the next pixels goes on, if the pixels doesn't match, and if the distance between pixels is greater than 10 then we draw a circle and there is the possibility of blockage. Although this technique gives the output for the input image being passed, but the output is not that accurate for all the image, this is because of the images that are been produced during angiogram videos (Fig. 3).



**Fig. 3.** Output for detection of possible blockage using OpenCV

A basic custom Convolutional Neural Network (CNN) was developed to detect coronary artery blockage on the coronary angiography images. We examined five models

with various architectures, network complexity and several weights: CNN, Res-Net, MobileNet, ResNet-150, and Inception ResNet. The light weight ResNet, and Inception ResNet with over 35 million weights, whereas customized CNN, ResNet and MobileNet with over 65 million weights were the most complex models selected for the study. Table 2 shows you the model training settings and Table 1 shows you the model characteristics (Fig. 4).

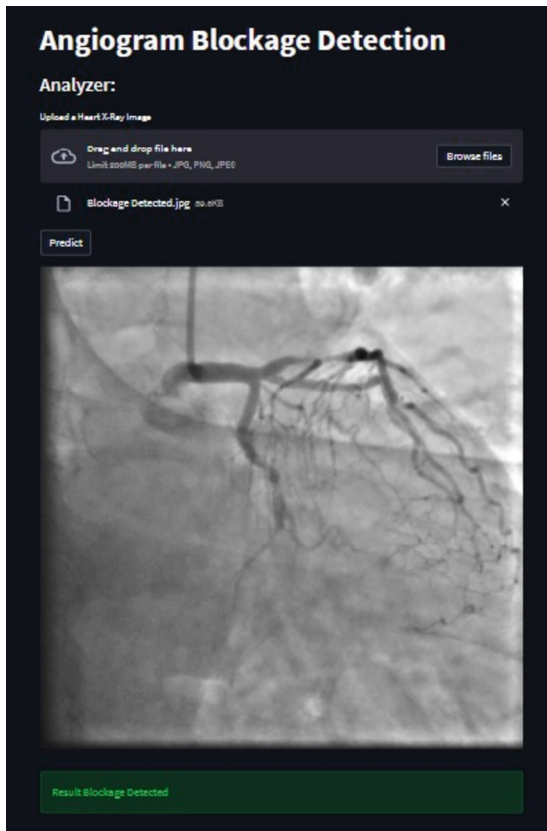
**Table 1.** Brief Characteristics of the models

Model	Inference time, ms (GPU)	Weights, min	Model size, mb
Custom CNN	38	65	412
Mobile Net V2	31	6.1	19
ResNet-50	30	25.6	114
Inception ResNet	34	55.9	241

**Table 2.** Model Operational Settings

Model	Input Size	Augmentation	Batch Size	LR
MobileNet V2	$128 \times 128 \times 3$	Random horizontal flip	32	0.04
Custom CNN	$128 \times 128 \times 3$	Random horizontal flip	32	0.04
ResNet-50	$128 \times 128 \times 3$	Random horizontal flip	8	0.003
Inception ResNet	$128 \times 128 \times 3$	Random horizontal flip	8	0.0003





**Fig. 4.** Web-based assistive diagnostic tool

## 5 Conclusion

This work presents the application of the OpenCV image processing framework for the segmentation of X-Ray angiograms and Deep learning models for the detection of blockage present or not using Customized CNN, MobileNet v2, ResNet-150 and Inception ResNet. Further the automated segmentation of vessels for stenosis detection using a deep neural net framework such as Region-based Convolutional Neural net, Inception Ney, Faster-RCNN, and MobileNet is presented.

The web application designed in Flask provides a user-friendly experience to obtain results from any type of data source – whether it is an image, pre-recorded video or live camera streaming. Once run on a system, the application can be accessed from any device on the network and due to its responsiveness, can also be accessed from a mobile device.

Integration of a Live Dashboard and Chat Bot within the application further increases its efficiency. The dashboard helps the user visualize the data concerning various parameters, hence enabling them to infer certain details. The Chat Bot makes it possible to fetch results from the database using simple queries in plain English, without the need for complex coding.

**Acknowledgments.** The Coronary X-Ray Angiogram dataset was acquired by cardiologist, SRMC at Sri Ramachandra Institute of Higher Education and Research, Tamil Nadu, India. The work reported herein was supported financially by the Universidad de Castilla-La Mancha (Spain) and the European Regional Development Funds (Reference: 2023-GRIN-34066).

## References

1. Klein, A.K., Lee, F., Amini, A.A.: Quantitative coronary angiography with deformable spline models. *IEEE Trans. Med. Imaging* **16**(5), 468–482 (1997). <https://doi.org/10.1109/42.640737>
2. Liu, I., Sun, Y.: Recursive tracking of vascular networking angiograms based on the detection-deletion scheme. *IEEE Trans. Med. Imaging* **12**(2), 334–341 (1993). <https://doi.org/10.1109/42.232264>
3. Quek, F.K.H., Kirbas, C.: Vessel extraction in medical images by wave-propagation and traceback. *IEEE Trans. Med. Imaging* **20**(2), 117–131 (2001). <https://doi.org/10.1109/42.913178>
4. Eiho, S., Qian, Y.: Detection of coronary artery tree using morphological operator. *Comput. Cardiol.* **24**, 525–528 (1997). <https://doi.org/10.1109/CIC.1997.647950>
5. Chen, W., Smith, R., Ji, S.Y., Ward, K.R., Najarian, K.: Automated ventricular systems segmentation in brain ct images by combining low-level segmentation and high-level template matching. *BMC Med Inform Decis Making.* **9**(1), 4 (2009). <https://doi.org/10.1186/1472-6947-9-S1-S4.11>
6. Wilkinson, M.H.F., Westenberg, M.A.: Shape preserving filament enhancement filtering. In: Niessen, W.J., Viergever, M.A. (eds.) *MICCAI 2001. LNCS*, vol. 2208, pp. 770–777. Springer, Heidelberg (2001). [https://doi.org/10.1007/3-540-45468-3\\_92](https://doi.org/10.1007/3-540-45468-3_92)
7. Iyer, K., et al.: <https://www.nature.com/articles/s41598-021-97355-8>
8. Lin, C.Y., Ching, Y.T.: Extraction of coronary arterial tree using cine X-ray angiograms. *Biomed. Eng. Appl. Basis Commun.* **17**, 111–120 (2005) <https://doi.org/10.4015/S1016237205000184>
9. Herrington, D.M., Siebes, M., Sokol, D.K., Siu, C.O., Walford, G.D.: Variability in measures of coronary lumen dimensions using quantitative coronary angiography. *J. Am. Coll. Cardiol.* **22**(4), 1068–1074 (1993). [https://doi.org/10.1016/0735-1097\(93\)90417-y](https://doi.org/10.1016/0735-1097(93)90417-y)
10. Canero, C., Radeva, P.: Vesselness enhancement diffusion. *Pattern Recogn. Lett.* **24**, 3141–3151 (2003). <https://doi.org/10.1016/j.patrec.2003.08.001>
11. Xia, S., et al.: Vessel segmentation of X-ray coronary angiographic image sequence. *IEEE Trans. Biomed. Eng.* **67**, 1338–1348 (2020). <https://doi.org/10.1109/2FTBME.2019.2936460>
12. Rawat, W., Wang, Z.: Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput.* **29**, 2352–2449 (2017). [https://doi.org/10.1162/2Fneco\\_a\\_00990](https://doi.org/10.1162/2Fneco_a_00990)
13. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 2481–2495 (2017). <https://doi.org/10.1109/TPAMI.2016.2644615>

14. Khanmohammadi, M., Engan, K., Sæland, C., Eftestøl, T. Larsen, A.I.: Automatic estimation of coronary blood flow velocity step 1 for developing a tool to diagnose patients with microvascular angina pectoris. *Front. Cardiovasc. Med.* **6**, 1 (2019). <https://doi.org/10.3389/fcvm.2019.00001>
15. Zhao, H., Shi, J., Qi, X., Wang, X. Jia, J.: Pyramid scene parsing network. In: 2017 IEEE Conference on Computer Vision Pattern Recognition. CVPR, pp. 6230–6239 (2017) <https://doi.org/10.1109/CVPR.2017.660>
16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)



# HIYAM, A New Moroccan Humanoid Robot for Healthcare Applications Using IoT and Big Data Analytics

Hiba Asri<sup>(✉)</sup> and Zahi Jarir

Computer Systems Engineering Laboratory, Computer Science Department, Faculty of Sciences  
Semlalia, Cadi Ayyad University, Marrakesh, Morocco  
hiba.asri@uca.ac.ma

**Abstract.** Advancements in Big Data, Robotics, Artificial Intelligence (AI), and the Internet of Things (IoT) have led to the emergence of humanoid robots, set to become commonplace in various aspects of our lives. Through machine learning algorithms and AI, machines can now learn from experiences and respond contextually. This has created a demand for real-time decision-making and assistance systems in healthcare. The paper proposes a new system composed of three components: an assisting subsystem, a decision-making subsystem, and the actors (patients and doctors). One integral part of this architecture is the HIYAM robot, a humanoid robot under development in Morocco. HIYAM is designed to learn, assist patients in medical facilities and homes, and make decisions based on different situations. The robot possesses diverse features such as conversing with patients, walking, searching for medications and definitions, entertaining patients, providing care, and detecting faces and emotions. Developing HIYAM is a complex endeavor that involves enhancing both the assisting and decision-making subsystems. Overall, this paper highlights the progress in humanoid robot technology, particularly the HIYAM robot, and its potential to revolutionize healthcare by assisting patients and supporting decision-making processes.

**Keywords:** AI · HIYAM · IoT · Healthcare · Assisting-System · Decision-Making

## 1 Introduction

Thanks to advancements in Big Data, artificial intelligence, and the Internet of Things (IoT), we notice the growing presence of humanoid robots in various fields [1, 2]. Unlike traditional robots used in industries like oil inspection and mine sweeping, humanoid robots possess the ability to learn from experience and adapt to different situations. They have the potential to assist people in everyday tasks and can be particularly beneficial in healthcare, education, and aging societies. Researchers propose a new Theory of Knowledge that recognizes intelligence, conscience, and principles as integral to humanoid robots, positioning them as successors to humans [3].

The integration of humanoid robots into daily life has sparked debates and curiosity in society. Providers are investing significant resources in creating models and architectures for everyday use. Market studies indicate a substantial increase in the demand for artificial intelligence humanoid robots, with the global personal robots market projected to reach \$34.1 billion by 2022 [3, 4]. It is also predicted that robots will replace 25% of human tasks by 2025 [5]. Research on humanoid robots dates back to 1966, with notable examples such as the WABIAN-2 robot developed by Waseda University and Honda's ASIMO robot.

The paper introduces a proposed architecture for a system where an artificial intelligence humanoid robot assists and makes real-time decisions. The system comprises three main parts: an assistant part that mimics human behavior, a decision-making subsystem that collects and analyzes data to make appropriate decisions, and a decision maker part that responds to abnormal situations to prevent undesired outcomes. This system has potential applications in education, healthcare, training, and other domains that require the integration of AI and predictive analytics [6, 7].

We focus on implementing the system in healthcare settings, specifically in hospitals where the humanoid robot serves as a nurse, assisting patients and making decisions in abnormal situations. Medical staff, such as doctors or nurses, play a role in reacting to information provided by the robot. In fact, communication between the assisting/Decision parts and doctors is made using a mobile application, which is a part that have been introduced in a previous work [8].

To implement the assisting part of the system, the authors introduce HIYAM, a Moroccan humanoid robot with various capabilities. HIYAM can engage in dialogue through speaking or chat interfaces, move in multiple directions, detect and recognize faces and emotions, read text from images, provide services such as reading books and sharing information about the current time and weather, define concepts and medications, conduct research on specific topics, and offer entertainment by telling jokes. The authors acknowledge that ongoing efforts are aimed at further improving HIYAM and the entire system.

The paper is structured into several sections, including a literature review on recent advancements in humanoid robotics. An overview of the proposed system's architecture is presented, highlighting the humanoid robot and decision-making subsystem. The case study focuses on the development of HIYAM, an artificial intelligence humanoid robot for healthcare applications, explaining its implementation and artificial intelligence capabilities. The paper concludes with future work plans.

## 2 Related Work

The creation of humanoid robots requires the collective efforts of researchers and specialists in various fields, including artificial intelligence, engineering, human-machine interaction, deep learning theories, and neuroscience. These fields have undergone significant advancements and revolutionary changes, enabling the development of humanoid robots. Current humanoid robots are capable of mimicking human behaviors such as walking, engaging in conversations, shaking hands, and analyzing facial expressions.

While we are in an era where humanoid robots can assist us in various activities and potentially replace humans in the near future, controlling them remains a challenge due

to their high degree of freedom. Therefore, control theory plays a crucial role in regulating and directing their actions. In the existing literature, several control methods for humanoid robots have been proposed and implemented during the development of such robots [9–12]: Virtual Model Control (VMC), Dynamic balance force control, Floating Body Inverse Dynamics (FBID), Inverse-kinematics force control (IKFC), Passivity-based balance control (PBBC).

In 2015, NASA developed a humanoid robot named Valkyrie that can operate in hazardous environments [14]. The robot was developed in collaboration with the Edinburgh Centre for Robotics to improve its performance. In addition, the Korean Advanced Institute of Science and Technology created HUBO2, a humanoid robot capable of walking and running in 2002.

Humanoid robots are not just designed for replacing humans in tasks, but also for entertainment purposes. For example, companies like De Agostini and Hanson Robotics have created robots like Robi and Sophia respectively for games and entertainment [15]. Robi is a DIY robot that buyers can build themselves, while Sophia is the first robot to have Saudi Arabian citizenship.

In Singapore, the Nadine robot has been developed by Nanyang Technological University and can express ersatz emotional responses through its 27 different facial and body movements [16]. In Tokyo, there is a humanoid robot that welcomes visitors to Japan, while a hotel in Nagasaki, Japan, employs a robot staff. The Henn'na hotel is staffed entirely by robots [17]. Additionally, researchers in Bangladesh have developed a humanoid robot named BONGO that can speak Bangla and English, detect faces, converse, and answer questions [17].

### 3 The Proposed System

Through this paper, we propose an architecture for building an artificial intelligence humanoid robot that can assist in healthcare decision-making in real-time. The robot has a human-like shape, moderate size and weight, and low power usage. The architecture is divided into three parts: 1) the assisting subsystem, 2) the decision-making subsystem, and 3) the involvement of doctors and nurses (see Fig. 1).

Some works of the two second parts of our proposed system, were already designed in a previous study where we collect data from sensors and then analyze them to get decision about miscarriage. But, in our current case, we will need adaptation for use in the new architecture because of the including of a third part which is the Humanoid Robot that use new sensors and IoT equipment; as well as developing a web application for hospital in addition to the mobile application that we created in our previous study. Unlike previous work, the proposed system is not about predicting outcome based on risk factors of a special disease, but it is designed for assisting patients, making decisions and help medical staff to take decision based on experience.

The robot is equipped with multiple tools to assist patients during treatment, such as conversing, explaining, and entertaining. It also collects healthcare data from sensors and analyzes it with Spark Big data tool to make real-time decisions, notify doctors, serve medicine, and give advice. The architecture involves the doctor or medical staff to take action in case of unpredictable situations. Communication between the assisting/decision parts and doctors is made using web and mobile applications.

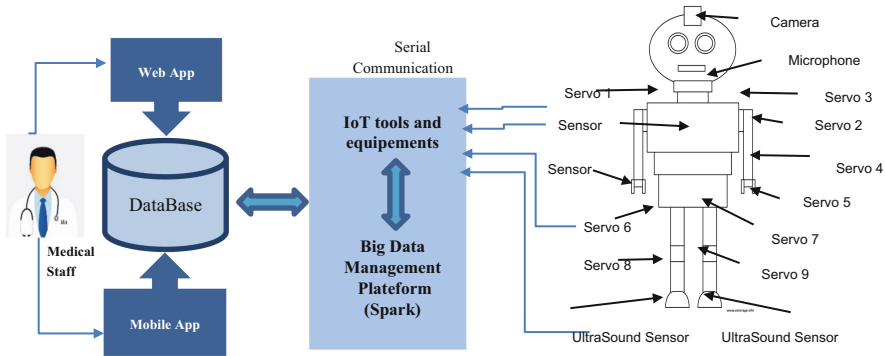


Fig. 1. Architecture of the proposed system

### 4 Case Study: “HIYAM”, A New Moroccan Artificial Intelligence Humanoid Robot for Healthcare Application

We suggest the construction of a humanoid robot named HIYAM to complement our proposed architecture. The purpose of this robot is to assist patients in hospitals, with the goal of improving people’s health and well-being and potentially saving lives. The HIYAM robot is a new development in Morocco that incorporates multiple functions for both assisting patients and making healthcare decisions. The creation of the robot is a challenging undertaking, and we are currently working to improve its design and functionality. Figure 2 and Fig. 3 provide an illustration of the proposed robot’s features and capabilities.

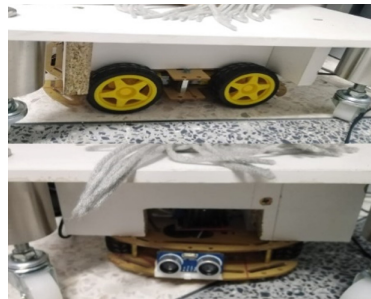


Fig. 2. HIYAM Humanoid robot (Real Picture)

Fig. 3. Chassis Robot of HIYAM

To facilitate the development of HIYAM, various tools and technologies from the fields of robotics, artificial intelligence, and predictive analytics were employed:

Raspberry Pi, Arduino Uno, healthcare sensors, servo motors, Movement kits ... among others.

As previously discussed, the proposed system includes a component focused on designing an artificial intelligence humanoid robot capable of providing assistance to humans and making real-time decisions to improve lives and promote well-being. The

development of HIYAM encompasses various functions pertaining to assistance and decision-making. Ongoing efforts are being made to further enhance the functions and capabilities of HIYAM for integration into the healthcare system, aiming to save lives, simplify tasks, and enhance overall health and quality of life for individuals. Table 1 summarizes some of the functions of HIYAM.

```

if total_results == 0:
    say("Sorry, there isn't any news about that topic ! Please choose another one.")
else:
    say("There are about " + str(
response.json()['totalResults']) + " different results, I will tell you the first one !")
    say("This article was written by : " + response.json()['articles'][0]['author'] + ". And titled : " +
response.json()['articles'][0]['title'] + " : " + response.json()['articles'][0][
'description'])

```

**Fig. 4.** Example of a general research

```

elif 'forward' in message:
    send_serial(b"U")
elif 'back' in message:
    send_serial(b"D")
elif 'left' in message:
    send_serial(b"L")
elif 'right' in message:
    send_serial(b"R")
elif 'stop' in message:
    send_serial(b"S")

```

**Fig. 5.** Type of direction of HIYAM's moves

## 5 Test and Evaluation

The proposed architecture still in its beginning. Therefore, validation and evaluation steps are required to asset the validity of the model. The validation of the prediction process is already evaluated in our previous study [8], but it needs to be adjust to the current study parameters and metrics.

In this work, we are working on testing HIYAM's functionalities (conversation, research, movement, detection, and recognition) in real world examples (see Fig. 4, Fig. 5, Fig. 6 and Fig. 7).

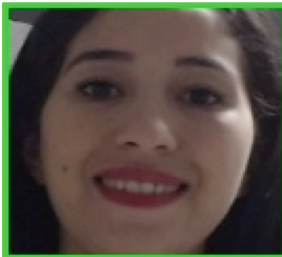
## 6 Conclusion and Future Works

This paper discusses the potential of humanoid robots in various fields, such as Big Data, IoT, and Artificial Intelligence. The paper proposes an architecture for a real-time assisting and decision-making system in healthcare, involving a humanoid robot named HIYAM and a decision-making component. HIYAM serves as a nurse robot, providing assistance and various services to patients. The decision-making component uses healthcare sensors and Spark Big Data tools for real-time analysis.



**Table 1.** Some Artificial Intelligence of Hiyam

Heading level	Example
Conversation	The system uses Google API to process human speech and find matching answers from a neural-based training dataset. It also retrieves information from Google and Wikipedia based on summarized questions. It employs a chatbot for conversations and allows for the addition of new conversations to the dataset. Its application in hospitals enables patients to have discussions and find comfort
Research	HIYAM can search for information or definitions using Wikipedia and the Google API. It provides search results and reads the first result aloud. In hospitals, patients can ask for medication definitions. HIYAM also performs news searches in various domains and presents the number of results before reading the first result
Movement	HIYAM, the humanoid robot, can move in different directions such as forward, backward, left, and right. It uses ultrasonic Distance Sensors to detect obstacles and adjusts its movement accordingly
Detection and Recognition	HIYAM captures faces using a camera connected to a Raspberry Pi and uses OpenCV and FER packages to detect motion and analyze emotions. The captured images are compared to a dataset, and HIYAM responds with the corresponding emotion, such as "You look Sad today."



**Fig. 6.** Face detected

```
img = plt.imread("test.jpg")
detector = FER(mtcnn=True)
text, score = detector.top_emotion(img) # Title Case looks Stun
font = cv2.FONT_HERSHEY_TRIPLEX
cv2.putText(im, str(text), (x + w, y), font, 1, (0, 0, 255), 2)
return text
```

**Fig. 7.** Face Emotion detected

The system aims to integrate the humanoid robot, decision-making subsystem, and medical professionals, while acknowledging the importance of human expertise in critical situations. Future plans include enhancing HIYAM’s capabilities, implementing it in hospitals, and involving doctors and nurses in the decision-making process. The paper presents a vision for an advanced humanoid robot system in healthcare, emphasizing real-time assistance and the collaboration between robots and human medical professionals.

## References

1. Asri, H., Jarir, Z.: Toward a smart health: big data analytics and IoT for real-time miscarriage prediction. *J. Big Data* **10**(1), 1–23 (2023). <https://doi.org/10.1186/s40537-023-00704-9>
2. Asri, H., Mousannif, H., Al Moatassime, H., Noel, T.: Big data in healthcare: challenges and opportunities. In: 2015 International Conference on Cloud Technologies and Applications (CloudTech), pp. 1–7 (2015). <https://doi.org/10.1109/CloudTech.2015.7337020>
3. Pistono, F.: Robots Will Steal Your Job, But That's OK: How to Survive the Economic Collapse and Be Happy. Federico Pistono
4. Breazeal, C., Dautenhahn, K., Kanda, T.: Social robotics. In: Siciliano, B., Khatib, O. (eds.) *Springer Handbook of Robotics*, pp. 1935–1972. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-32552-1\\_72](https://doi.org/10.1007/978-3-319-32552-1_72)
5. Alesich, S., Rigby, M.: Gendered robots: implications for our humanoid future. *IEEE Technol. Soc. Mag.* **36**(2), 50–59 (2017). <https://doi.org/10.1109/MTS.2017.2696598>
6. Asri, H.: Big data and IoT for real-time miscarriage prediction a clustering comparative study. *Procedia Comput. Sci.* **191**, 200–206 (2021). <https://doi.org/10.1016/j.procs.2021.07.025>
7. Asri, H., Mousannif, H., Al Moatassime, H., Zahir, J.: Big data and reality mining in healthcare: promise and potential. In: El Moataz, A., Mammass, D., Mansouri, A., Nouboud, F. (eds.) *ICISP 2020. LNCS*, vol. 12119, pp. 122–129. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-51935-3\\_13](https://doi.org/10.1007/978-3-030-51935-3_13)
8. Asri, H., Mousannif, H., Al Moatassime, H.: Reality mining and predictive analytics for building smart applications. *J. Big Data* **6**(1), 66 (2019). <https://doi.org/10.1186/s40537-019-0227-y>
9. Kahraman, C., Deveci, M., Boltürk, E., Türk, S.: Fuzzy controlled humanoid robots: a literature review. *Robot. Auton. Syst.* **134**, 103643 (2020). <https://doi.org/10.1016/j.robot.2020.103643>
10. Righetti, L., Buchli, J., Mistry, M., Schaal, S.: Inverse dynamics control of floating-base robots with external constraints: a unified view. In: 2011 IEEE International Conference on Robotics and Automation, pp. 1085–1090 (2011). <https://doi.org/10.1109/ICRA.2011.5980156>
11. Al-Gallaf, E.A.: Multi-fingered robot hand optimal task force distribution: neural inverse kinematics approach. *Robot. Auton. Syst.* **54**(1), 34–51 (2006). <https://doi.org/10.1016/j.robot.2005.09.016>
12. Bhattacharya, S., Luo, A., Dutta, S., Miura-Mattauch, M., Mattauch, H.J.: surface recognition and speed adjustment of humanoid robot using external control circuit. In: 2020 International Symposium on Devices, Circuits and Systems (ISDCS), pp. 1–4 (2020). <https://doi.org/10.1109/ISDCS49393.2020.9263013>
13. Van Pinxteren, M.M., Wetzels, R.W., Rüger, J., Pluymaekers, M., Wetzels, M.: Trust in humanoid robots: implications for services marketing. *J. Serv. Market.* **33**(4), 507–518 (2019). <https://doi.org/10.1108/JSM-01-2018-0045>
14. Radford, N.A., et al.: Valkyrie: NASA's First bipedal humanoid robot. *J. Field Robot.* **32**(3), 397–419 (2015). <https://doi.org/10.1002/rob.21560>
15. Retto, J.: SOPHIA, FIRST CITIZEN ROBOT OF THE WORLD (2017)
16. Thalmann, N.M., Tian, Li., Yao, F.: Nadine: a social robot that can localize objects and grasp them in a human way. In: Prabaharan, S.R.S., Thalmann, N.M., Kanchana Bhaaskaran, V.S. (eds.) *Frontiers in Electronic Technologies. LNEE*, vol. 433, pp. 1–23. Springer, Singapore (2017). [https://doi.org/10.1007/978-981-10-4235-5\\_1](https://doi.org/10.1007/978-981-10-4235-5_1)
17. Ahsanul Sarkar Akib, A.S.M., Farhan Ferdous, M., Biswas, M., Khondokar, H.M.: Artificial intelligence humanoid BONGO robot in Bangladesh. In: 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), pp. 1–6 (2019). <https://doi.org/10.1109/ICASERT.2019.8934748>



# Optimal Forecast Combination with Univariate Models for Natural Gas Prices in Spain

Roberto Morales-Arsenal , María Pilar Zazpe-Quintana ,  
and Jesús María Pinar-Pérez  

CUNEF Universidad, Madrid, Spain  
[jesusmaria.pinar@cunef.edu](mailto:jesusmaria.pinar@cunef.edu)

**Abstract.** The forecast combination for energy prices has generally focused on the combination of neural network algorithms. This paper extends the literature by combining conventional models with neural network models. This work analyzes the point forecast combination applied to the daily prices of natural gas in Spain during 2023. To do this, a model with a deterministic trend and ARIMA process in the residuals (RMSE = 1.40) is combined with a neural networks model (RMSE = 1.65) offering excellent results in terms of predictive ability (RMSE = 0.88) in the short term. The best forecasting performance was obtained by applying the Bates and Granger combination method, estimating very similar weights to the two estimated models: 0.58 and 0.42, respectively. The presented combined model outperforms the models reported in the literature. The result of the forecast exercise indicates that the fall in the price of natural gas will continue.

**Keywords:** Energy prices · daily frequency · point forecast combination

## 1 Introduction

Despite the introduction of numerous renewable energy sources in recent years, fossil fuel-based energy sources (oil, coal, and natural gas) still account for the majority of total energy consumption. Among the fossil fuels, natural gas is the cleanest of all in terms of pollutant emissions. Natural gas became commercially available at the end of the 19th century, and since then, gas consumption per capita has increased dramatically, reaching 5,000 kWh per capita in 2019 [1]. According to the International Energy Agency (IEA), the use of natural gas accounted for almost a third of the total growth in energy demand over the past decade and is expected to continue to grow strongly in the coming years. Despite this greater growth in demand for gas, its price has fallen substantially in the

---

Point forecast combination with simple models and advanced models with high frequency data.

last year for two fundamental reasons: 1) an average increase in temperatures (climate change) 2) due to the inflation reduction law approved and, in particular in Spain, due to the so-called *Iberian exception* whose main objective is to control energy prices. This drop in gas prices has affected energy prices in general, allowing inflation to moderate. This has allowed central banks not to implement more restrictive monetary policies that would affect economic growth. In any case, the analysis of the evolution of natural gas prices, as one of the main sources of energy, is a key factor for decision-makers: Central Banks (to design the monetary policy), investors and households.

According to [2] it is possible to distinguish three stages: first stage from 1972 to 1992 characterized by the use of traditional methods of time series analysis such as the Box-Jenkins methodology [3] a second stage from 1992 to 2006 is stage of AI-based models [4–6] and finally, the last stage, from 2006 to present, is characterized by the appearance of deep learning models, the combination of algorithms and the return to the traditional models [7, 8].

The previous studies about forecast combination have focused on the combination of algorithms, mainly from machine learning techniques. This study tries to fill the gap in the research by combining the predictions of traditional models with more advanced machine learning models such as neural networks. A gas price index has been selected with a daily frequency obtained in an organized gas market such as the Iberian gas market (Mibgas<sup>1</sup>) in Spain during the year 2023, which has allowed us to make good predictions in the short term. The comparison between the methods used individually and the combination of both has made it possible to capture both linear (with traditional methods) and non-linear (by means of neural networks) relationships, which are reflected in lower mean squared errors and higher r-squared.

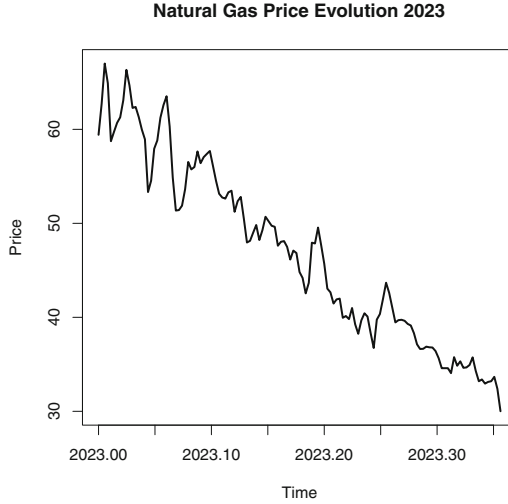
This paper is structured as follows. Section two describe the data set used. Section three contains the forecasting strategy and the empirical results. The last section is devoted to the final conclusions.

## 2 Data Description

The data set is composed by 131 daily observations obtained from Mibgas [9] from 01/01/2023 to 05/11/2023. In particular, the time series represents MIBgas-Es index (Eur/KWh). This is the average weighted price of all the trades arranged of all the products with delivery on the same Gas day at the PVB (point virtual balance) in all the trading sessions that have already concluded. The forecasting exercise was carried out ten periods ahead ( $h = 10$ ). Figure 1 depicts the time series for natural gas price taking into account in this work.

---

<sup>1</sup> Mibgas is part of the European Gas Market Model (Gas Target Model).



**Fig. 1.** Natural Gas Prices for 2023

The time series shows the following characteristics:

1. Downward trend.
2. No seasonal component is observed.
3. No existence of outliers or aberrant data values.
4. Linear relationship between time and price.
5. It is a non-stationary process. The mean change over time but the variance seems to be constant except at the beginning of the sample.

### 3 Forecasting Techniques and Results

Two forecasting approaches are used: 1) traditional model such as **ARIMA** models using the Box-Jenkins methodology. These models have an excellent predictive ability in the short term. This kind of models have been combined with the inclusion of deterministic components (trends); and 2) Neural networks for time series forecasting. In particular, with feed-forward neural networks with a single hidden layer and lagged inputs for forecasting univariate time series [10].

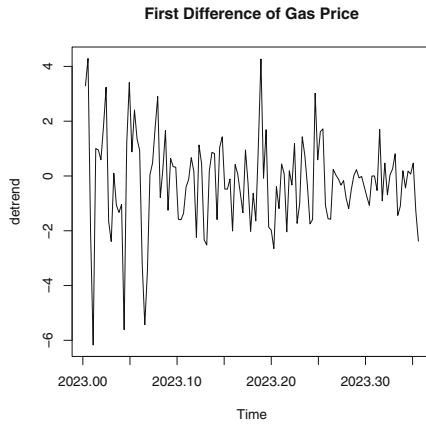
#### 3.1 Box-Jenkins Methodology

A very well-known methodology is the so called *Box-Jenkins methodology* (1974). This approach is based on the information of the past values of the time series. This considers that the time series has been generated by a stochastic process (PGD) assuming that the components trend and cycle are stochastic in

nature. The most general ARIMA model (multiplicative model  $ARIMA(p, d, q) \times ARIMA(P, D, Q)$ ) would be the following:

$$\Phi_P(B^s)\Phi_p(B)\nabla_s^D\nabla^d\text{Ln}(y_t) = \mu + \Theta_Q(B^s)\Theta_q(B)a_t \tag{1}$$

where  $\nabla$  is the number of differences (regular and seasonal differences) which we must apply to transform the series into stationary. As previously mentioned, the price time series does not show a pattern in the seasonal lags.  $\text{Ln}$  denotes the neperian logarithm. This methodology is based on the following steps: 1) identification of the ARMA process (once the time series is stationary), 2) Estimation, 3) Diagnosis and 4) Forecasting.



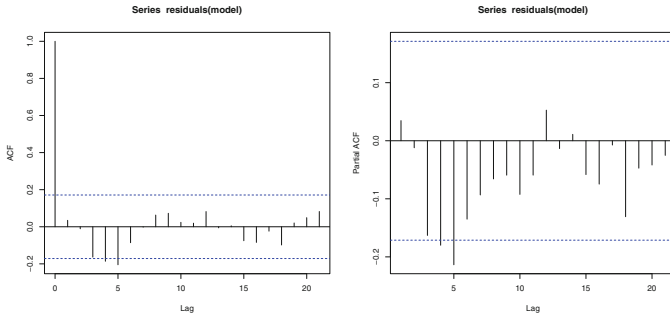
**Fig. 2.** First difference of Natural Gas Price

A test of unit roots was carried out, in particular, the Kwiatkowski-Phillips-Schmidt-Shin test figures out if a time series is stationary around a mean or linear trend, or is non-stationary due to a unit root. The null hypothesis for the test is that the data is stationary. The KPPS unit root test shows that the times series is not stationary (p-value = 0.001). Therefore, a regular difference is necessary. Figure 2 shows the first difference of the time series.

Once the times series is stationary, a stochastic process of type ARIMA (0,1,2) is identified. Note that different ARIMA structures were tested in the seasonal part and no results were obtained. Moreover, no seasonal structure was observed in the correlograms.

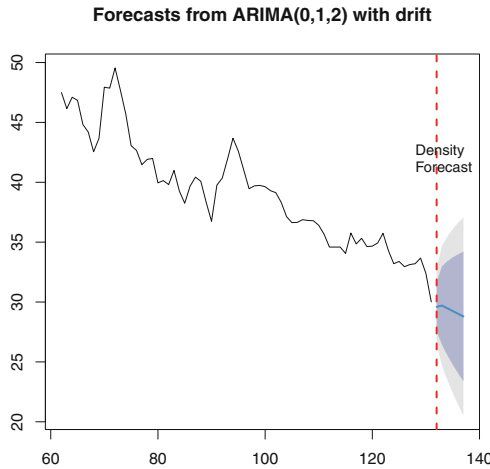
$$\nabla\text{Price}_t = -0.22 + a_t + \underset{(0.11)}{0.24} a_{t-1} + \underset{(0.13)}{0.19} a_{t-2} \quad a_t \sim N(0, \sigma^2 = 2.55)$$

The plots of the residuals show that they are white noise (zero mean and constant variance). The first value of the autocorrelation function (ACF) does not count since it is always 1 (see Fig. 3).



**Fig. 3.** Model Diagnosis

Finally, once the model has been validated it can be used for the prediction stage. Figure 4 shows the point forecast (solid blue line) and the density forecast of this model at two levels of confidence: 80% (dark grey area) and 95% (light grey area).



**Fig. 4.** Forecast from ARIMA model

### 3.2 Deterministic Trend + ARMA model

The possibility of combining the *deterministic trend* plus an ARMA process for the error term has been considered.

$$y_t = \beta_0 + \beta_1 time + \varepsilon_t \tag{2}$$

where  $\varepsilon_t$  follows an ARMA process. This model was compared with an alternative model with stochastic trend (taking a difference, very similar to a random walk)

plus an ARMA process. In this case, deterministic modeling is more suitable in terms of predictive ability.

The estimated model is as follows:

$$\begin{aligned}
 y_t &= 63.10 - 0.24time + \varepsilon_t \\
 \varepsilon_t &= 1.04\varepsilon_{t-1} - 0.43\varepsilon_{t-2} + a_t \\
 a_t &\sim N(0, 2.036)
 \end{aligned}$$

An ARMA(2,0,0) process with drift was identified in the residuals of the model where  $a_t$  is a white noise.

The Fig. 5 shows the density forecasting for this model.

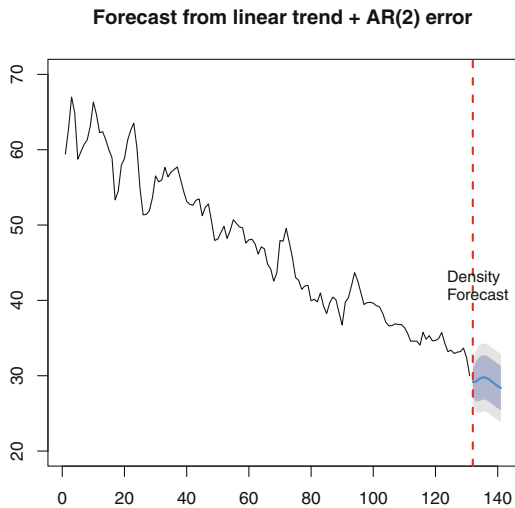


Fig. 5. Forecast Deterministic trend + ARMA model

### 3.3 Artificial Neural Networks

One of the most used methods in machine learning for forecasting are the so-called artificial neural networks (ANN). Neural networks use nonlinear algorithms to both predict and classify. Although they have been criticized on many occasions as black box methods, nevertheless, neural networks have obtained excellent results in forecasting applications. The nnetar function in R was used.

$$y_t = f(y_{t-1}) + \varepsilon_t \tag{3}$$

where  $y_{t-1} = (y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4})$  is vector containing the lagged values of natural gas price and  $f$  denotes the neural network with four hidden nodes and a single layer. Figure 6 depicts the density forecast.



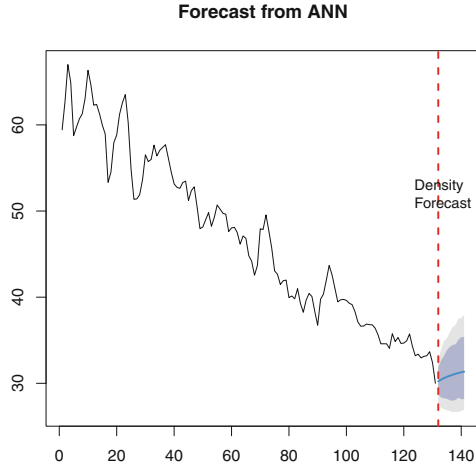


Fig. 6. Forecast from ANN

According to Table 1 the model with the best performance (lowest RMSE) is the combination between a deterministic linear trend plus ARMA process.

Table 1. Summary model performance

Model	Forecast ability			Goodness of fit statistics		
	RMSE	MAPE	MASE	AIC	BIC	$R^2$
Linear trend	2.25	3.73	0.21	590.66	599.28	0.94
Linear trend + ARIMA (2,0,0)	1.40	2.19	0.85	472.02	486.39	0.97
ARIMA (0,1,2)	1.57	2.45	0.96	495.94	507.41	0.97
ANN(1,1)	1.65	2.49	0.98	NA	NA	NA

### 3.4 Forecast Combination

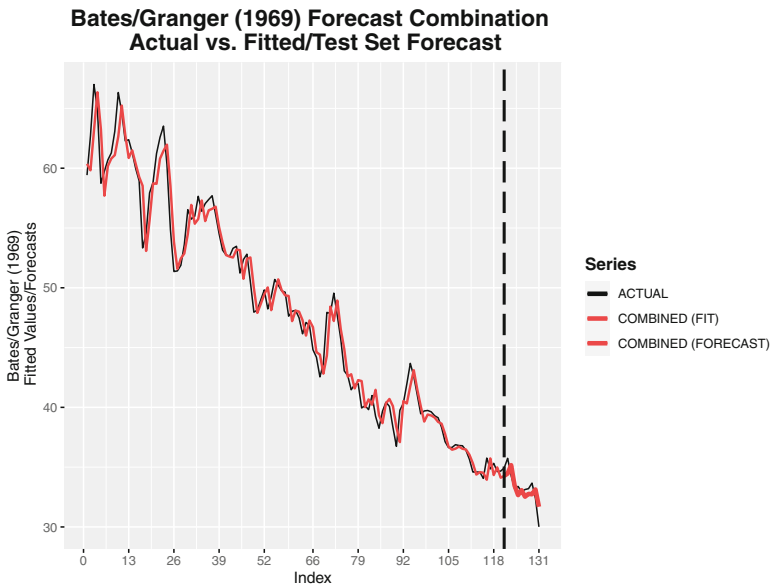
Based on the seminal work of Bates and Granger, empirical studies have shown that forecasting combination improves the performance of individual models either because it is capable of capturing different data characteristics or because it uses different information sets. In this work, two ways of combining individual predictions have been used: 1) *à la Granger* [11,12], this approach builds on portfolio diversification theory and uses the diagonal elements of the estimated mean squared prediction error matrix in order to compute combination weights and 2) OLS [13]. The OLS combination method uses ordinary least squares to estimate the weights with a regression model.

The predictions of the linear trend plus ARIMA model ( $m1$ ) have been combined together with the neural network ( $m2$ ) predictions. Table 2 gives a summary of forecast combination.

**Table 2.** Summary forecast combination

Model	Forecast Combination			Combination weights	
	RMSE	MAE	MAPE	$weight_{m1}$	$weight_{m2}$
Bates-Granger	0.88	0.76	2.31	0.58	0.42
OLSR	0.92	0.81	2.43	0.95	0.05

The results show that the Granger combination (with very similar weights) outperforms (lowest RMSE) the OLS method. Figure 7 shows the adjusted values and the point forecast of the combination.



**Fig. 7.** Bates-Granger Forecast Combination

## 4 Conclusions

The use of forecast combination for natural gas research (demand, supply or prices) has focused almost entirely on combining very complex machine learning algorithms in order to improve individual performance. This paper shows how the combination of a simple (conventional) model that captures a large proportion of the variability of the time series (linear behaviour) together with a neural network model that captures non-linear relationships obtains excellent results in the short term ( $h = 10$ ) for daily price data. The presented combined model (RMSE = 0.88) outperforms the models reported in the literature. The weights




of the Granger combination indicate a slightly higher weight of the conventional model (0.58) on the neural network. The result of the forecast combination exercise indicates that the fall in the price of natural gas will continue.

## References

1. Doruk, S., Hamurcuoglu, K.I., Ersoy, M.Z., Murat-Tunc, K.M., Günay, M.E.: Forecasting long-term world annual natural gas production by machine learning. *Resour. Policy* **80**, 103224 (2023)
2. Liu, J., Wang, S., Wei, N., Chen, X., Xie, H., Wang, J.: Natural gas consumption forecasting: a discussion on forecasting history and future challenges. *J. Nat. Gas Sci. Eng.* **90**, 103930 (2021)
3. Conejo, A.J., Plazas, M.A., Espinola, R., Molina, A.B.: Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. *IEEE Trans. Power Syst.* **20**(2), 1035–1042 (2005)
4. Pelikan, E., Simunek, M.: Risk management of the natural gas consumption using genetic algorithms. *Neural Netw. World* **15**(5), 425–436 (2005)
5. Szoplik, J.: Forecasting of natural gas consumption with artificial neural networks. *Energy* **85**, 208–220 (2015)
6. Khotanzad, A., Elragal, H., Lu, T.L.: Combination of artificial neural-network forecasters for prediction of natural gas consumption. *IEEE Trans. Neural Networks* **11**(2), 464–473 (2000)
7. Laib, O., Khadir, M.T., Mihaylova, L.: Toward efficient energy systems based on natural gas consumption prediction with LSTM recurrent neural networks. *Energy* **177**, 530–542 (2019)
8. Malik, H., Yadav, A.K., Márquez, F.P.G., Pinar-Pérez, J.M.: Novel application of relief algorithm in cascaded artificial neural network to predict wind speed for wind power resource assessment in India. *Energ. Strat. Rev.* **41**, 100864 (2022)
9. Mibgas Homepage. <http://www.mibgas.es>. Accessed 4 May 2023
10. Hyndman, R.J., Athanasopoulos, G.: *Forecasting Principles and Practice*, 2nd edn. OTexts, Australia (2018)
11. Bates, J.M., Granger, C.W.: The combination of forecasts. *J. Oper. Res. Soc.* **20**(4), 451–468 (1969)
12. Timmermann, A.: Forecast combinations. In: Elliott, G., Timmermann, A. (eds.) *Handbook of economic forecasting*, vol. 1, pp. 135–196 (2013)
13. Granger, C.W., Ramanathan, R.: Improved methods of combining forecasts. *J. Forecast.* **3**(2), 197–204 (1984)



# CNN for Efficient Objects Classification with Embedded Vector Fields

Oluwaseyi Igbasanmi , Nikolay M. Sirakov  <sup>(✉)</sup>, and Adam Bowden 

Texas A&M University-Commerce, Commerce, TX 75428, USA

Nikolay.Sirakov@tamuc.edu

**Abstract.** Classification methods use image object features to distinguish between objects and assign them to classes. In the present study we develop a convolutional neural network (CNN) optimized to classify images with embedded vector fields (VFs), generated on the solution  $\hat{u}(x, y)$  of the Poisson equation, which contains the image function in its right-hand side. The embedded VF features subject to extraction, by our CNN, are trajectories and singular points (SP), which augment the image object features. The aim of this paper is to validate that the set of augmented image features increases the separability of the image objects and improves the classification statistics. To reach the aim, we implement our CNN along with four contemporary CNNs to classify two public image databases COIL100 and ISIC2020 as well as their derivatives with embedded VFs. The obtained results are presented in the paper and confirm that embedding VFs with real and complex SPs increases the classification statistics.

**Keywords:** Vector Fields · Singul Points · Machine Learning · Classification

## 1 Introduction

In recent years, vector fields (VFs) found applications in computer vision and machine learning (ML). For example, the authors in [6] created a benchmark VF dataset useful for any local deep learning features extraction. Further, they designed a neural network (NN) to extract steady frames from non-steady VF. A general algorithm for the description of a flow field is outlined in [7]. This paper develops functions to describe a VF in the vicinity of its SPs, and an algorithm that recognizes the linear combinations of basic vector flows by estimating, with a NN, the projection coefficients. In paper [8] is developed a deep learning convolutional NN (CNN) to determine the boundary vortices. The training data for the new CNN is generated by using parametric VF. The integration of the Clifford product in a NN results in Clifford correlation and convolution to be used for template matching on VFs [9]. In [10] the authors introduce a novel approach for comparing projections obtained from multiple stages of a NN and visualize differences in data perception. Then the differences in projections are

transformed into VF trajectories that represent the flow of information. It produces layouts that highlight abstract structures identified by NNs. A method that extracts SPs from dense motion fields is developed in [11], which considers the presentation of a VF as a sum of irrotational and solenoidal components (solutions of the Poisson equation) that lead to a complex potential form of the VF, which yields to the Polya's equation form of the VF [12].

In image classification, features like color, texture, shape characterize image objects. The following papers show the application of VFs for objects characterization as well. Using the solution of the Poisson equation the method in [15] generated a scalar field, applied to capture objects features with high complexity. In [16] the authors discussed the local algebras of linear VF's that could be used in the modeling of physical space by building the dynamic flows of VFs on 8-dimensional cylindrical or toroidal manifolds. In [17] is proposed VFs simplification through successively removing pairs of connected SPs. A disadvantage of this method is that it relies on a stable extraction of the connections but it's numerical integration may be unstable. In [18], the authors built a novel simplification algorithm based on degree theory derived from the recently introduced topological notion of robustness. Paper [13] defines and implements Clifford Convolution on VF for matching patterns.

As it is known the NNs possess non-linear mapping capabilities which are invariant according to translations, scaling, and rotations. Therefore, the NNs would be capable of recognizing VF characteristics if they are invariant according to the above affine transformations [14,21]. Thus, the authors of [14] developed a CNN to efficiently classify VF features, mainly singular points (SPs). To train the CNN they generated approximately 1500 training images of SP shapes.

Paper [19] connects VFs and image objects' features. The authors developed two gradient VFs with real valued SPs and showed their correlation with image objects and their parts. Further, [20] applied a gradient VF to guide an active contour evolution to segment multiple image objects. In [21] the authors developed two VFs with real and complex SPs and proved that the VF SPs shapes are invariant according to scaling, translation, and weak rotation. This property implies that the VF SPs are useful for image classification and embedding them into images will increase the classification statistics of supervised ML methods. To validate this advantage, we develop a CNN optimized to classify images with embedded VFs. Further, we implement our CNN along with four contemporary CNNs on the image databases COIL100 [22] and ISIC2020 [23], their derivatives with embedded VFs, designed in [21], and the databases covered with Gaussian noise and VFs. The results are presented in Tables 3 to 9.

The paper is organized as follows. Section 2 presents VFs' basics; Sect. 3 develops a new CNN optimized to classify images with embedded VFs and describes the image databases COIL100 [22] and ISIC2020 [23]; Sect. 4 presents experiments with five CNN classifiers on the two original, the original datasets with embedded VFs, and the two original image datasets covered with Gaussian noise and embedded VFs; the paper ends with listing the conclusions, the contribution and the advantages of this study.

## 2 VF Basic Notions and Statements

**Definition 1:** A vector field is described as a function  $H$  that assigns a vector  $H(x, y) = (V(x, y), W(x, y))$ ,  $V, W \in \mathbb{R}$  to every point  $(x, y) \in \mathbb{R}^2$ .

**Definition 2:** Given a VF  $H(x, y) = W(x, y)i + V(x, y)j$  defined on the image frame  $\Omega \subset \mathbb{R}^2$ , the points  $z = xi + yj$  where  $H(x, y) = (0, 0)$  are its singularities.

The VF’s trajectories in the vicinity of a SP are called shape or pattern of the SP. The different types of patterns are springing, sink, saddle, spiral in, spiral out and center (see [20,21] for SPs shape’s description). The first three patterns are known as real shaped SPs while the rest are the complex shaped SP.

The Polya’s model [12] of a non gradient VF (GVF) is given as [21]

$$\bar{H}(z) \approx \bar{H}(0) + \frac{z}{2}(\nabla \cdot \bar{H}) + \frac{iz}{2}(\nabla \times \bar{H}) + c\bar{z}. \tag{1}$$

Since the curl of a GVF is 0, then  $\frac{iz}{2}(\nabla \times \bar{H}) = 0$  and its Polya’s representation is described as [12,21]

$$H(z) \approx H(0) + \frac{z}{2}(\nabla \cdot \bar{H}) + c\bar{z}. \tag{2}$$

Consider the two VFs with real and complex shaped SPs described in [21]

$$v_{\hat{u}} = \hat{u}_x i - \hat{u}_y j, \text{ and } v_{\hat{\phi}} = \hat{\phi}_x i - \hat{\phi}_y j. \tag{3}$$

In [21], we proved that  $v_{\hat{u}}$  and  $v_{\hat{\phi}}$  possess real and complex shapes around the vicinity of their SPs. Also, the shapes of these SPs are invariant according to scaling with the same coefficients on both axes; translation; rotation with angle  $\alpha$  such that  $\alpha \in [-\frac{\pi}{4}, \frac{\pi}{4}] \cup \{\pi\} \cup \{2\pi\}$  [21]. This property states that the SPs’ shapes of VFs represent an image object feature which may increase the classification statistics of supervised ML methods if a VF is embedded into an image. To validate this advantage, we develop a CNN optimized to detect, extract and recognize image and VF shapes and features.

## 3 Design of a VF Optimized CNN

CNNs are NNs that apply image convolution and pooling layers for image recognition and classification. Consider an image  $I(x, y)$  with sizes  $M \times N$  and a filter  $\mu(x, y)$  with sizes  $m \times n$ , with  $n = 2p + 1, m = 2k + 1$ . The convolution slides  $\mu(x, y)$  over  $I(x, y)$  and at each pixel calculates the formula and replaces the pixel  $I(x, y)$  with the result of Eq. (4)

$$I \circ \mu = \sum_{j=-k}^k \sum_{i=-p}^p \mu(i, j)I(x - i, y - j). \tag{4}$$

EfficientNet B0 [1] is a well-established deep NN. Its basic building block is the Mobile Inverted Bottleneck Conv Block, MBconv [2], which is very fast and makes accurate predictions. The main components of the EfficientNet B0 [1] architecture are listed in Table 1. It has 237 layers trained on the ImageNet database. The model is calibrated several times by adjusting the hyper-parameters in the last layer. The main concept of the EfficientNet B0 architecture is to start from one high-quality yet compact baseline model presented in Table 1 and progressively scale each of its dimensions systematically with a mixed set of scaling coefficients.

**Table 1.** EfficientNet B0 network: architecture [1].

Stage	Operator	Resolution	#channels	#layers
1	Conv3 × 3	224 × 224	32	1
2	MBCConv1, k3 × 3	112 × 112	16	1
3	MBCConv6, k3 × 3	112 × 112	24	2
4	MBCConv6, k5 × 5	56 × 56	40	2
5	MBCConv6, k3 × 3	28 × 28	80	3
6	MBCConv6, k5 × 5	14 × 14	112	3
7	MBCConv6, k5 × 5	14 × 14	192	4
8	MBCConv6, k3 × 3	7 × 7	320	1
9	Conv1 × 1/Pooling/FC	7 × 7	1280	1

### 3.1 Proposed CNN Model

We developed our new CNN architecture by updating the EfficientNet B0 replacing its last layer with a fully connected layer (FC), and other blocks. These blocks include Batch normalization (BN), dropout, and pooling layer, where Conv means Convolution. The BN operation constrains the output of the layer in a specific range, forcing zero mean and standard deviation one. That works as a regularization, increasing the stability of the NN and accelerating the training. The Dropout operation also acts as a regularization by blocking few neurons and thus emulating a bagged ensemble of multiple NNs, for each mini-batch on training. The dropout parameter defines the number of blocked neurons (from 0 to 100% of the neurons of a layer). To accelerate the convergence of the loss function to zero we employed the ReLu activation function, which is formally defined as  $f(x) = \max(0, x)$ . We named the new designed CNN SeyNet.

The SeyNet model aims to detect and use the image objects' features along with the shapes of the embedded VF SPs to boost classification accuracy. For this purpose we, constructed the model with a focus primarily placed on the number of layers and feature maps and simultaneously considering the efficacy in terms of accuracy and other classification metrics of the model, which is a fully connected (dense) CNN. Its architecture is described in Table 2.

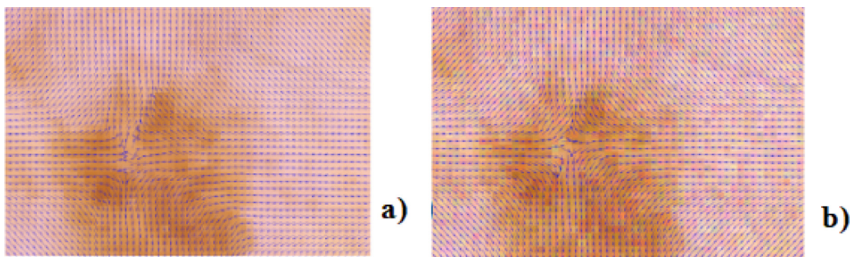
**Table 2.** SeyNet architecture using EfficientNet B0 [1] as a baseline model.

Stage	Operator	Resolution	#channels	#layers
1–9	EfficientNet B0	$300 \times 300$	32	1
10	BN/Dropout	$7 \times 7$	1280	1
11	Pooling	$2 \times 2$	512	1
12	FC/BN/ReLU	1	512	1
13	FC/Softmax	1	2	1

### 3.2 Datasets and Hyper-parameters

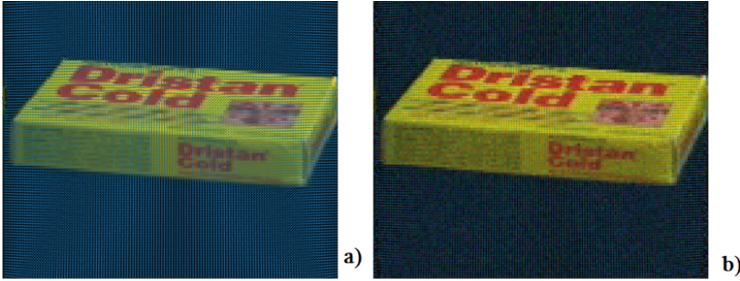
To validate the above aim, we run experiments with the following datasets.

1. ISIC 2020 [23] dataset contains 33,126 labeled RGB color training images. The dataset consists of cancerous and non-cancerous skin lesions images from over 2,000 patients. Using ISIC 2020 training images [23] we created four new image databases available at: <https://inside.tamuc.edu/academics/colleges/scienceEngineeringAgriculture/departments/mathematics/research/augmented-features-repository/default.aspx>. The first pair is generated by embedding the VFs  $v_{\hat{u}}$  and  $v_{\hat{\phi}}$ . Hence, we named the new databases  $v_{\hat{u}}(I)$  and  $v_{\hat{\phi}}(I)$  respectively. The second pair is generated by adding Gaussian noise to every ISIC2020 training image. Then we embedded the VFs  $v_{\hat{u}}$  and  $v_{\hat{\phi}}$  and named the new image databases  $v_{\hat{u}}(I+N)$  and  $v_{\hat{\phi}}(I+N)$  respectively. Figure 1 shows images of size  $246 \times 168$  from the image databases  $v_{\hat{u}}(I)$  and  $v_{\hat{u}}(I+N)$ .
2. COIL100 dataset has 7200 RGB color images [22], which contain 100 distinct objects that are rotated through  $360^\circ$  with a step  $5^\circ$  per rotation. Hence COIL100 has 100 classes and 72 objects per class. Using COIL100 we developed four new image datasets. The first pair is created by embedding the VFs  $v_{\hat{u}}$  and  $v_{\hat{\phi}}$  and the new datasets are named  $v_{\hat{u}}(C)$  and  $v_{\hat{\phi}}(C)$  respectively. The second pair is generated by adding Gaussian noise to every COIL100 image. Then we embedded the VFs  $v_{\hat{u}}$  and  $v_{\hat{\phi}}$  and named the new image datasets  $v_{\hat{u}}(C+N)$  and  $v_{\hat{\phi}}(C+N)$  respectively. Figure 2 shows images of size  $420 \times 420$  from the datasets  $v_{\hat{u}}(C)$  and  $v_{\hat{u}}(C+N)$ .



**Fig. 1.** A skin lesion from the image database: **a)**  $v_{\hat{u}}(I)$ ; **b)**  $v_{\hat{u}}(I+N)$ .





**Fig. 2.** An object from the image database: **a)**  $v_{\hat{u}}(C)$ ; **b)**  $v_{\hat{u}}(C + N)$ .

The SeyNet CNN model was implemented using the module Keras in the Tensorflow library. Since the parameters of interest during training are the weights and biases, the model’s optimizer is the Keras built-in Adam Optimizer with the hyper-parameters: learning rate=0.001,  $\mu_1 = 0.9$ ,  $\mu_2 = 0.999$ ,  $\epsilon = 1e - 08$  and decay=0.0. The model is trained with the number of epochs set to 30 or until there is no change in the loss metric. To combat over-fitting of the model, the loss metric is monitored using a method known as Early-Stopping.

## 4 Experimental Results

We coded in Python the designed SeyNet architecture with EfficientNet B0 as a base network for our experiments. SeyNet is optimized for classifying images with embedded VFs. To validate the concept that embedding VFs in images augments their features and boosts classification metrics we conducted experiments with the, described above, ten image databases implementing SeyNet and four contemporary CNNs.

**Table 3.** Classification results by SeyNet and other contemporary models.

ISIC 2020			
Model	Training ACC (%)	Validation ACC (%)	Testing ACC (%)
Dense Net121 [3]	95.78	85.12	83.29
Res Net50 [4]	95.77	82.79	83.81
Inception ResNet V2 [24]	96.33	93.59	86.05
EfficientNet B0 [1]	<b>98.62</b>	93.01	<b>90.07</b>
SeyNet	97.27	<b>93.69</b>	86.67
COIL 100			
Dense Net121 [3]	96.15	85.44	85.84
Res Net50 [4]	96.14	83.10	86.38
Inception ResNet V2 [24]	96.70	93.94	88.68
EfficientNet B0 [1]	<b>98.77</b>	93.36	<b>92.83</b>
SeyNet	97.65	<b>94.04</b>	89.41

**Five Classifiers on the Original Image Datasets.** We first applied SeyNet and four contemporary CNNs to classify the original databases ISIC2020 and COIL100. The results are shown in Table 3. To classify the former database we randomly select 2500 skin lesion images from its training set. Out of them 2000 are used for training and 500 for testing. The former set is split to 310 malignant and 1690 benign. For testing images, 110 are the malignant and 390 are benign. The above settings were used by all five CNNs.

The best training accuracy (AC), 98.62%, and testing AC, 90.07% are achieved by EfficientNet B0. SeyNet has the highest validation AC of 93.69% and the second-best training and testing AC of 97.27% and 86.67% respectively. Note that, the training loss converged after the 25th epoch for SeyNet.

In the case of COIL100, for all five classifiers we used 55 images per class for training and the remaining 17 for testing. The results are given in Table 3. The highest training 98.77% and testing 92.83% ACs were obtained by EfficientNet B0 while SeyNet showed the highest validation AC of 94.04%.

We ran each of the five classifiers five times on each dataset ISIC2020 and COIL100. Then we calculated the mean of every set of five runs and reported the results in Table 3.

**Table 4.** Testing results for SeyNet on all datasets. The highest results are in bold.

Dataset	AC(%)	PR(%)	SN(%)	SP(%)
ISIC 2020	86.67	81.11	76.04	91.67
$v_{\hat{u}}(I)$	90.33	<b>83.33</b>	84.26	<b>93.33</b>
$v_{\hat{u}}(I + N)$	<b>92.09</b>	83.00	<b>92.22</b>	91.90
$v_{\hat{\phi}}(I)$	84.33	77.78	72.16	90.14
$v_{\hat{\phi}}(I + N)$	85.26	71.26	85.26	85.26
COIL100	89.41	69.84	89.40	89.41
$v_{\hat{u}}(C)$	91.27	74.85	91.27	91.27
$v_{\hat{u}}(C + N)$	<b>92.97</b>	80.00	<b>92.98</b>	<b>92.97</b>
$v_{\hat{\phi}}(C)$	90.08	71.57	90.09	90.08
$v_{\hat{\phi}}(C + N)$	90.55	<b>82.84</b>	90.55	90.56

**SeyNet on the ten Image Datasets.** Table 4 shows the classification results of SeyNet, which we ran five times on each of the ten image databases (described in Sect. 3.2) and calculated the mean of each set of five runs. The experiments were conducted under the same settings and ratios as those used for the experiments shown in Table 3. Also, the images were rescaled to sizes  $441 \times 294$ . Table 4 illustrates that the database  $v_{\hat{u}}(I + N)$  has the best-noted AC of 92.09%. It is 5.42% increase according to the SeyNet AC on the original ISIC2020, and 2.02% increase according to the best results on ISIC2020 obtained by EfficientNet B0 in Table 3. The precision (PR), sensitivity (SN), and specificity (SP) are computed at 83.33%, 92.22%, and 93.33%, respectively. Further, the classification AC of  $v_{\hat{u}}(C + N)$  92.97% shows an improvement of 0.14% according to the highest AC of COIL100 classification by EfficientNet B0, and 3.56% improvement according to

the SeyNet AC. Thus, Table 4 validates the statement that images with embedded VFs increase the classification statistics. Moreover, adding Gaussian noise and embedding VF further increases the values.

**The Five Classifiers on the Entire Original Datasets.** Table 5 illustrates the classification of the original image datasets COIL 100 [22] and ISIC 2020 [5] with the five classifiers. For the ISIC2020 classification we used the entire dataset of training images. The model was trained using 2484 images selected randomly from the entire set of 33126 training images. Of these 2484 images, 2000 are benign while 484 are malignant. For testing, the remaining 30642 training images were used. Of these images, 100 are malignant, while the rest are benign. The best AC of 91.69% and SP - 91.69% are achieved with EfficientNet B0. SeyNet has the next best AC - 90.10% and a SP - 90.10%. Inception ResNet V2 has the best SN - 93%. The PR value is low for all the models due to the very high ratio of benign to malignant images for testing.

To classify COIL100, we used 10 images per class for training and 62 for testing. The best AC of 90.99% is achieved with EfficientNet B0, the highest PR of 70.46% by ResNet 50, while the top SP 91% and SN 91.54% were obtained by Inception ResNet V2. The slightly lower results if Table 5 are due to the fact that in this case 10 images were used for training, while for the results in Tables 3 and 4, 55 images were used for training.

Note that every result presented in Table 5, as well as Tables 6, 7, 8 and 9, is the mean of two runs under same conditions and hyper-parameters.

**Table 5.** SeyNet and other contemporary models on original datasets.

ISIC 2020				
Model	AC	PR	SN	SP
DenseNet 121 [3]	89.40	2.76	92.00	89.39
ResNet 50 [4]	90.20	2.98	92.00	90.19
Inception ResNet V2 [24]	89.60	2.84	<b>93.00</b>	89.59
EfficientNet B0 [1]	<b>91.69</b>	<b>3.50</b>	92.00	<b>91.69</b>
SeyNet	90.10	2.89	90.00	90.10
COIL 100				
DenseNet 121 [3]	88.72	69.19	90.55	88.71
ResNet 50 [4]	89.51	<b>70.46</b>	91.11	89.51
Inception ResNet V2 [24]	88.91	68.11	<b>91.54</b>	<b>91.00</b>
EfficientNet B0 [1]	<b>90.99</b>	69.09	91.43	88.91
SeyNet	89.41	69.84	89.40	89.41

**The Five Classifiers on  $v_{\hat{u}(I)}$ ,  $v_{\hat{u}(C)}$  and  $v_{\hat{\phi}(I)}$ ,  $v_{\hat{\phi}(C)}$  Datasets.** Recall, the first pair of datasets is created from ISIC2020 and COIL100 by embedding into their images the VF  $v_{\hat{u}}$ . The upper part of Table 6 shows the classification of  $v_{\hat{u}(I)}$  dataset with the five classifiers which used the same settings as those

used to produce the results in Table 5. In the upper part of Table 6 the top AC of 93.24% is obtained by SeyNet and it is an increase of 1.56% according to the highest AC by EfficientNet B0 in the upper part of Table 5. Moreover Table 6 assures that all classifiers, except Dense Net 121, improved their classification statistics when classified  $v_{\hat{u}}(I)$ . The highest AC 91.85% with  $v_{\hat{u}}(C)$  is obtained by EfficientNet B0 and shows an increase of 0.86% according to the highest classification of COIL100 in Table 5. The second pair of datasets  $v_{\hat{\phi}}(I)$ ,  $v_{\hat{\phi}}(C)$  is created from ISIC2020 and COIL100 by embedding into their images the VF  $v_{\hat{\phi}}$ . The datasets are classified with the five CNNs (Table 7) under the same set ups as those used for ISIC2020, COIL100 (Table 5 ) and  $v_{\hat{u}}(I)$ ,  $v_{\hat{u}}(C)$  (Table 6) databases. Studying Table 7 one may observe that its classification results are lower if compared to those in Tables 5 and 6. An exception are the results for SeyNet on  $v_{\hat{\phi}}(C)$  which are the highest in Table 7 and higher then those of SeyNet in Table 5, which still confirms the claim that embedding VF into images boosts their classification.

**Table 6.** SeyNet and the five contemporary NNs on  $v_{\hat{u}}$  embedded datasets.

$v_{\hat{u}}(I)$				
Model	AC	PR	SN	SP
DenseNet 121	88.88	2.64	92.00	88.87
ResNet 50	91.19	3.38	94.00	91.19
Inception ResNet V2	92.50	3.86	92.00	92.50
EfficientNet B0	92.90	4.19	95.00	92.89
SeyNet	<b>93.25</b>	<b>4.45</b>	<b>96.00</b>	<b>93.20</b>
$v_{\hat{u}}(C)$				
Model	AC	PR	SN	SP
DenseNet 121	87.88	72.58	90.04	88.04
ResNet 50	90.16	73.31	92.00	90.34
Inception ResNet V2	91.45	70.78	90.04	91.64
EfficientNet B0	<b>91.85</b>	<b>75.10</b>	<b>92.98</b>	<b>92.03</b>
SeyNet	91.27	74.85	91.27	91.27

**The Five Classifiers on  $v_{\hat{u}}(I+N)$ ,  $v_{\hat{u}}(C+N)$  and  $v_{\hat{\phi}}(I+N)$ ,  $v_{\hat{\phi}}(C+N)$  Datasets.** The following Tables 8 and 9 show the classification of the four image databases generated from ISIC2020 and COIL100, which we covered with Gaussian noise and then embedded the VFs  $v_{\hat{u}}$  and  $v_{\hat{\phi}}$ . For these experiments we use 2000 images for training and 500 for testing to classify  $v_{\hat{u}}(I+N)$  and  $v_{\hat{\phi}}(I+N)$  datasets as well as 55 images per class for training and 17 for testing to classify  $v_{\hat{u}}(C+N)$ ,  $v_{\hat{\phi}}(C+N)$  datasets. Note that, these settings are the same as those used for Tables 3 and 4. The classification of  $v_{\hat{u}}(I+N)$  is reported in the upper part of Table 8 and show an increase of the AC for all classifiers if compared with Table 3. The highest AC 92.27% comes from EfficientNet B0, while the highest SN 92.22% and SP of 91.90% come from SeyNet. The classification of  $v_{\hat{u}}(C+N)$

**Table 7.** SeyNet and other contemporary models on  $v_{\hat{\phi}}$  covered datasets.

$v_{\hat{\phi}}(I)$				
Model	AC	PR	SN	SP
DenseNet 121	73.50	1.01	<b>83.00</b>	73.47
ResNet 50	74.34	1.00	79.00	74.32
Inception ResNet V2	74.79	1.02	80.00	74.78
EfficientNet B0	72.50	0.97	82.00	72.47
SeyNet	<b>78.19</b>	<b>1.22</b>	82.00	<b>78.18</b>
$v_{\hat{\phi}}(C)$				
Model	AC	PR	SN	SP
DenseNet 121	84.82	70.20	87.79	84.79
ResNet 50	85.79	68.94	87.17	85.77
Inception ResNet V2	86.31	70.04	89.33	86.30
EfficientNet B0	83.67	69.96	87.63	83.64
SeyNet	<b>90.08</b>	<b>71.57</b>	<b>90.09</b>	<b>90.08</b>

is shown in the lower part of Table 8 and demonstrates AC increase for all classifiers except EfficientNet B0. The highest AC 92.97% comes from SeyNet which shows an increase of 3.56% if compared to Table 3. The classification results of  $v_{\hat{\phi}}(I+N)$  are shown in Tables 9. The upper part represents an increase in the AC of the upper two classifiers while the next three decrease their AC if compare to Table 3. The lower part of Tables 9 represents an increase of all classifiers except EfficientNet B0. The top AC 90.85% come from DenseNet 121 and shows an increase of 5.11% if compared to Table 3.

**Table 8.** SeyNet and other contemporary models on  $v_{\hat{u}}$  and noise covered datasets.

$v_{\hat{u}}(I + N)$				
Model	AC	PR	SN	SP
DenseNet 121	89.88	79.27	87.23	89.67
ResNet 50	88.25	<b>85.58</b>	85.55	88.28
Inception ResNet V2	89.89	64.52	89.29	90.50
EfficientNet B0	<b>92.27</b>	83.02	91.68	91.67
SeyNet	92.09	83.00	<b>92.22</b>	<b>91.90</b>
$v_{\hat{u}}(C + N)$				
Model	AC	PR	SN	SP
DenseNet 121	90.05	86.40	89.05	90.04
ResNet 50	89.57	82.49	89.24	89.34
Inception ResNet V2	90.15	87.62	90.08	90.11
EfficientNet B0	90.61	<b>89.89</b>	89.97	90.24
SeyNet	<b>92.97</b>	80.00	<b>92.98</b>	<b>92.97</b>

**Table 9.** SeyNet and five contemporary models on  $v_{\hat{\phi}}$  and noise covered datasets.

$v_{\hat{\phi}}(I + N)$				
Model	AC	PR	SN	SP
DenseNet 121	84.60	72.05	84.15	84.60
ResNet 50	85.35	<b>73.48</b>	84.35	84.34
Inception ResNet V2	85.79	73.03	85.79	85.79
EfficientNet B0	<b>86.76</b>	72.30	<b>86.75</b>	<b>86.75</b>
SeyNet	85.26	71.26	85.26	85.26
$v_{\hat{\phi}}(C + N)$				
Model	AC	PR	SN	SP
DenseNet 121	<b>90.85</b>	86.47	<b>90.84</b>	<b>90.85</b>
ResNet 50	89.65	87.25	89.66	89.65
Inception ResNet V2	90.05	87.69	90.05	90.05
EfficientNet B0	90.15	<b>88.69</b>	90.15	90.15
SeyNet	90.55	82.84	90.55	90.56

Note that every result presented in Tables 5, 6, 7, 8 and 9, is the mean of two runs under same conditions and hyper-parameters.

## 5 Conclusion

The main contribution of this paper is that we validated the statement formulated in [21], which states that embedding a VF into images increases the classification statistics of supervised ML methods. For the purpose of this validation, we developed a CNN (SeyNet) and used the two public datasets ISIC2020 [23] and COIL100 [22]. Then, we embedded into them the VFs  $v_{\hat{u}}$  and  $v_{\hat{\phi}}$  and generated the four additional image datasets  $v_{\hat{u}}(I)$ ,  $v_{\hat{u}}(C)$  and  $v_{\hat{\phi}}(I)$ ,  $v_{\hat{\phi}}(C)$ . Further, we added Gaussian noise to the original datasets and embedded the above VFs to generate four datasets named  $v_{\hat{u}}(I + N)$ ,  $v_{\hat{u}}(C + N)$  and  $v_{\hat{\phi}}(I + N)$ ,  $v_{\hat{\phi}}(C + N)$ . Then we classified the ten image datasets with SeyNet and four contemporary CNNs.

Comparing Tables 5 and 6, one may observe that the classification of databases with embedded VF  $v_{\hat{\phi}}$  increased in AC as well as PR, SE, and SP, with the exception of DenseNet 12. It clearly validates the advantage of embedding VF SPs in images. The very small PR values are due to the very small ratio between malignant and benign testing images (100 malignant versus greater than 30,000 benign).

Comparing Table 5 with Table 7 shows that embedding  $v_{\hat{\phi}}$  into COIL100 leads to increased SeyNet classification measures AC, PR, SE, and SP. The statistics of the remaining four classifiers declined. This drop in the classification statistics was even higher for ISIC2020. This deficiency occurs because VF  $v_{\hat{\phi}}$  has a far

larger set of SPs distributed on the objects boundaries. This situation is worse for skin lesions because of their color and texture entropy. The conclusion is that embedding  $v_{\hat{\phi}}$  increases classification statistics in images with less detail.

Consider now Tables 3 and 8 where the experiments were conducted under the same ratios and hyper-parameters. Note, Table 3 contains results from classifying the original databases, while Table 8 show results from the original datasets, where Gaussian noise is populated, then  $v_{\hat{u}}$  is embedded. One can tell that, except for EfficientNet B0, all classifiers significantly increased their statistics. Adding Gaussian noise increased the classification statistics of the image databases with embedded  $v_{\hat{\phi}}$ , but only SeyNet exceeded its statistics in classifying the original COIL100. As well, DenseNet 12 and ResNet 50 exceeded their ISIC2020 classification statistics.

One may conclude from the above discussion that adding Gaussian noise further increases classification if  $v_{\hat{u}}$  is embedded afterwards. The same holds for  $v_{\hat{\phi}}$ , but only for objects with less detail. Even in this case, the increased statistics could be mixed compared with original database classification statistics.

Future work would improve upon the above foundation through further studying the reasons for AC decline. Also, we will develop an approach for training data augmentation with images where VF is embedded.

## References


1. Putra, T.A., Rufaida, S.I., Leu, J.S.: Enhanced skin condition prediction through machine learning using dynamic training and testing augmentation. *IEEE Access* **8**, 40536–40546 (2020). <https://doi.org/10.1109/ACCESS.2020.2976045>
2. Zhao, J., Zhang, Y., He, X., Xie, P.: COVID-CT-dataset: A CT scan dataset about COVID-19, CoRR, vol. abs/2003.13865, (2020). [arXiv: 2003.13865](https://arxiv.org/abs/2003.13865). [Online]. Available: <https://arxiv.org/abs/2003.13865>
3. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269 (2017). <https://doi.org/10.1109/CVPR.2017.243>
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770–778 (2016). <https://doi.org/10.1109/CVPR.2016.90>
5. Rotemberg, V., et al.: A patient-centric dataset of images and metadata for identifying melanomas using clinical context. *Sci. Data* **8**, 34 (2021). <https://doi.org/10.1038/s41597-021-00815-z>
6. Kim, B., Günther, T.: Robust reference frame extraction from unsteady 2D vector fields with convolutional neural networks. In: Gleicher, M., Leitte, H., Viola, I. (Guest Eds.) *Eurographics Conference on Visualization (EuroVis) 2019 V. 38, N 3*, Computer Graphics Forum 2019, The Eurographics Association and John Wiley & Sons Ltd. (2019.) [arXiv:1903.10255v1](https://arxiv.org/abs/1903.10255v1)
7. Branca, A., Attolico, G., Stella, E., Distante, A.: Classification and segmentation of vector flow fields using a neural network. *J. Mach. Vis. Appl.* **10**, 174–187 (1997)
8. Berenjkoub, M., Chen, G., Gunther, T.: Vortex boundary identification using convolutional neural network. In: *2020 IEEE Visualization Conference (VIS) 2020*, pp. 261–265 (2020). <https://doi.org/10.1109/VIS47514.2020.00059>

9. Ebling, J., Scheuermann, G.: Template matching on vector fields using Clifford algebra. In: Gurlebeck, K., Konke, C. (eds.) 17th International Conference on the Application of CS and Math in Architecture and Civil Engineering, Weimar, Germany, 12–14 July (2006)
10. Cantareira, G.D., Etemad, E., Paulovich, F.V.: Exploring neural network hidden layer activity using vector fields. *Information* **11**(9), 426 (2020). <https://doi.org/10.3390/info11090426>
11. Corpetti, T., Memin, E., Perez, P.: Extraction of singular points from dense motion fields: an analytic approach. *J. Math. Imaging Vis.* **19**, 175–198 (2003)
12. Lucian-Miti, I., Liliana, C.: Classification of holomorphic functions as pólya vector fields via differential geometry. *Mathematics* **9**, 1890 (2021). <https://doi.org/10.3390/math9161890>
13. Ebling, J., Scheuermann, G.: Clifford convolution and pattern matching on vector fields. *IEEE Visual.* **2003**, 193–200 (2003). <https://doi.org/10.1109/VISUAL.2003.1250372>
14. Bina, T., Yib, L.: CNN-based flow field feature visualization method. *Int. J. Performability Eng.* **14**(3), 434–444 (2018). <https://doi.org/10.23940/ijpe.18.03.p4.434444>
15. Tari, S., Genctav, M.: From a non-local Ambrosio-Tortorelli phase field to a randomized part hierarchy tree. *J. Math. Imaging Vis.* **49**(1), 69–86 (2014)
16. Bayak, I.: Applications of the local algebras of vector fields to the modelling of physical phenomena, vol. 38 (2015). <https://doi.org/10.7546/jgsp-38-2015-1-23>
17. Zhang, E., Mischaikow, K., Turk, G.: Vector field design on surfaces. *ACM Trans. Graphics* **25**(4), 1294–1326 (2006). <https://doi.org/10.1145/1183287.1183290>
18. Skraba, P., Wang, B., Chen, G., Rosen, P.: 2D vector field simplification based on robustness. In: *IEEE Pacific Visualization Symposium 2014*, pp. 49–56 (2014). <https://doi.org/10.1109/PacificVis.2014.17>
19. Chen, M., Sirakov, N.M.: Poisson equation solution and its gradient vector field to geometric features detection. In: Fagan, D., Martin-Vide, C., O’Neill, M., Vega-Rodriguez, M.A. (eds.) *Theory and Practice of Natural Computing. Lecture Notes in Computer Science()*, vol. 11324, pp. 36–48. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-04070-3\\_3](https://doi.org/10.1007/978-3-030-04070-3_3)
20. Bowden, A., Sirakov, N.M.: Active contour directed by the Poisson gradient vector field and edge tracking. *J. Math. Imaging Vis.* **63**, 665–680 (2021). <https://doi.org/10.1007/s10851-021-01017-3>
21. Igbasami, O., Sirakov, N.M.: On the usefulness of the vector field singular points shapes for classification. *Iter. J. Appl. Comput. Math.* (2024). Accepted for publication. <https://www.researchsquare.com/article/rs-2862010/v1>
22. Nene, S.A., et al., Columbia object image library(coil-100), Technical Report CUCS-006-96 (1996)
23. International Skin Imaging Collaboration. SIIM-ISIC 2020 Challenge Dataset. International Skin Imaging Collaboration (2020). <https://doi.org/10.34970/2020-ds01>
24. Szegedy, C., Ioffe, S., Vanhoucke, V.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI-17)*, pp. 4278–4284 (2017)





# CEIMVEN: An Approach of Cutting Edge Implementation of Modified Versions of EfficientNet (V1-V2) Architecture for Breast Cancer Detection and Classification from Ultrasound Images

Sheekar Banerjee<sup>1,2</sup>  and Md. Kamrul Hasan Monir<sup>1</sup>

<sup>1</sup> Department of Artificial Intelligence and Machine Learning, KaleidoSoft, Zagreb, Croatia  
sheekar.banerjee@gmail.com

<sup>2</sup> IUBAT- International University of Business Agriculture and Technology, Dhaka 1230, Bangladesh

**Abstract.** Undoubtedly breast cancer identifies itself as one of the most widespread and terrifying cancers across the globe. Millions of women are getting affected each year from it. Breast cancer remains the major one for being the reason of largest number of demise of women. In the recent time of research, Medical Image Computing and Processing has been playing a significant role for detecting and classifying breast cancers from ultrasound images and mammograms, along with the celestial touch of deep neural networks. In this research, we focused mostly on our rigorous implementations and iterative result analysis of different cutting-edge modified versions of EfficientNet architectures namely EfficientNet-V1 (b0-b7) and EfficientNet-V2 (b0-b3) with ultrasound image, named as CEIMVEN. We utilized transfer learning approach here for using the pre-trained models of EfficientNet versions. We activated the hyper-parameter tuning procedures, added fully connected layers, discarded the unprecedented outliers and recorded the accuracy results from our custom modified EfficientNet architectures. Our deep learning model training approach was related to both identifying the cancer affected areas with region of interest (ROI) techniques and multiple classifications (benign, malignant and normal). The approximate testing accuracies we got from the modified versions of EfficientNet-V1 (b0- 99.15%, b1- 98.58%, b2- 98.43%, b3- 98.01%, b4- 98.86%, b5- 97.72%, b6- 97.72%, b7- 98.72%) and EfficientNet-V2 (b0- 99.29%, b1- 99.01%, b2- 98.72%, b3- 99.43%) are showing very bright future and strong potentials of deep learning approach for the successful detection and classification of breast cancers from the ultrasound images at a very early stage.

**Keywords:** Deep Learning · Neural Networks · Image Processing · Computer Vision · Transfer Learning · Medical Image Computing

## 1 Introduction

Deep learning techniques are being used in a wide range of industries and have demonstrated impressive computational performance in areas such as speech recognition, natural language processing, video analysis, image processing, computation, analysis and classification. Additionally, deep neural networks combine many layer types, including FC (fully connected) layers, convolutional neural networks (CNN), and recurrent neural networks (RNN) [1]. Multi-level deep neural networks (DNNs) are used to build diverse neural networks that can recognize and categorize traits from an input of huge unlabeled training data. CNNs appear to be the neural model that processes and analyzes images the best. We are getting closer at extracting features from photographs that encompass the many elements of the underlying problem as a result of CNN's performance. Since DCNN's convolution and subsampling layers may automatically extract visual features from a given patch, they do not require data-focused and appropriately decomposed cores. Big data processing demands a sophisticated infrastructure and combination of powerful devices. The processing of enormous amounts of data has been made easier by advancements in hardware processing units, which have also sped up research on DCNN. A graphics processing unit (GPU) with CUDA support performs applications 10 to 100 times faster than a central processor unit (CPU) with hundreds of compute cores. By using parallel computing, the typical classification difficulties for medical images can be effectively decreased. As a result, the machine learning approach needs simultaneous Deep Neural Network models using CPU and GPU.

Breast Cancer appears to be as most prevalent cancer in the world for women. The World Health Organization (WHO) reports that the leading causes of mortality worldwide in 2018 were severe diseases like heart disease and breast cancer. According to statistics provided by WHO, 17.7 million people worldwide lose their lives to cardiovascular illnesses (CVDs) each year, accounting for 31% of all fatalities worldwide. Heart attacks and strokes near to the end of their account for 80% of deaths [2].

It is a known fact that compared to other types of cancer, breast cancer has a disproportionately high mortality rate in both developing and wealthy nations. Although the causes of breast cancer remain mostly unknown, extensive research and investigation have shown that the disease occurs with some frequency. Hereditary family history, obesity or inadequate exercise, alcohol and tobacco addiction, exposure to ionizing radiation, having children too late or not at all, and not producing enough milk to feed the young are high-risk factors. The woman will experience symptoms of this disease as a result of these factors. For the treatment of breast cancer and increasing survival, early detection is crucial. Patients should receive an early diagnosis even when medical resources are limited in order to perform better. This can be done by being aware of early signs and symptoms and promptly referring them to therapy [3]. The risks of death as a result are going down because early-stage barriers are being effectively removed and access to medical services for breast cancer is being improved. Mammography, clinical biopsy, and routine manual self-check are just a few of the techniques that make up screening and have been studied as breast cancer assessment instruments. The most well-known of them is mammography, which records breast features using low-energy X-rays and generates the corresponding medical images for additional diagnosis. In comparison to other optional methods, mammography is thought to be one of the top contributions to

lowering mortality and treating breast cancer because of its high efficacy. Cardio CTG using Doppler ultrasound monitoring, among other similar methods, is used for the early diagnosis of cardiac problems [4].

In our proposed research, we majorly concentrated on our exact implementations and iterative result analysis of several cutting-edge modified EfficientNet architectures, specifically EfficientNet-V1 (b0-b7) and EfficientNet-V2 (b0-b3) with ultrasound image. We named it as CEIMVEN (Cutting-Edge Implementation with Modified Efficient-Nets). Here, we used a transfer learning strategy to make use of the pre-trained EfficientNet versions models. In our specially modified EfficientNet architectures, we turned on the hyper-parameter tuning procedures, added completely linked layers, eliminated the rare outliers, and recorded the accuracy results. We used multiple classifications (benign, malignant and normal) and region of interest (ROI) algorithms to identify the cancer-affected areas in our deep learning model training strategy. The main goal of this research was to detect and classify breast cancer in real time from ultrasound images using more recent and cutting-edge deep neural network architectures. We sought to dispel the myth that improved accuracy may be obtained from using mammograms in medical image processing. We aimed to extract the best prediction accuracies from supposedly outdated ultrasound images by embedding cutting-edge neural networks into them. Unquestionably, we completed a really difficult and brave study project, and the outcomes were incredibly rich. The modification of EfficientNet-V1 (b0-b7) and EfficientNet-V2 (b0-b3) using a unique DL model and ultrasound images has never been attempted in prior studies.

Following the rest of this paper is: a literature review describing the previous research in the field of breast cancer detection and classification using other neural architectures, computational techniques, algorithms, datasets, optimizers; our proposed methodology of CEIMVEN implementation; experimental result analysis of our method. Finally, the conclusion and future research aspects originated from our research effort has been spotlighted.

## 2 Related Works

The medical imaging community has been conducting frequent observations in recent years thanks to the development of effective and potent CNN architectures. We reviewed the most related previous research works which we deeply admired but at the same time we implied to make improvements of too.

Experiments were conducted as a computer-generated diagnosis method for classification relying on both binary classification and multi-classification. They used a combined approach of neural networks (DNNs) like Res-Net-18, Shuffle-Net and Inception-Net-V3 by approaching transfer learning on the publicly available dataset of “BreakHis”. For Res-Net, Inception-Net-V3, and Shuffle-Net, respectively, their method produced the best average accuracy for binary classification of benign or malignant cancer cases of 99.7%, 97.66%, and 96.94%. ResNet, Inception-Net-V3, and Shuffle-Net each had multi-class classification accuracy averages of 97.81%, 96.07%, and 95.79%, respectively [5]. A new multi-layered CNN architecture had been introduced for clinical mammograms with better cross-folding and accuracy more than 95% [6].

A synthesized image modelling using GAN-CNN-wavelet architecture with MIAS dataset has shown quite a good amount of accuracy (approximately 87%) previously [7]. In the study of "MultiNet", the authors showed an implementation of a concept related to transfer. The image analysis were conducted by using pre-trained models, including VGG16, DenseNet-201 and NasNet-Mobile, where their estimation of accuracy was 98% [8]. Four mammography imaging datasets—normal, benign, and malignant—were employed in a study, with the basic classifiers being a variety of deep CNN models, including Inception V4, ResNet-164, VGG-11, and DenseNet121. The Gompertz function was used to provide fuzzy rankings of the fundamental classification procedures as part of their ensemble approach. Particularly noteworthy was the 99.32% accuracy rate of their Inception V4 ensemble model using fuzzy rank-based Gompertz function [9]. In order to prepare computed tomography (CT), a study has proposed an automated segmentation model based on traditional CNN-based deep neural networks for the breast cancer CT. Their model had three steps that functioned in a cascade, making it applicable to actual situations with roughly 80% accuracy [10]. By developing a multi-input classification model that took advantage of convolutional neural networks' strengths in image analysis, breast cancer detection using thermal images from various angles was made possible. The Database for Research with Infrared Image, the most extensively known public database of breast thermal pictures, was utilised in their application of our technique. Their top model had an accuracy of 97%, a ROC curve area of 0.99, and an 83% sensitivity [11]. Using an amassed dataset, a deep learning method was introduced for categorizing hematoxylin-eosin-stained breast cancer microscopy images into normal tissue, benign lesion, in situ carcinoma, and invasive carcinoma. This method utilized the Xception's six intermediate layers. This model had a kappa value of 0.965 and a precision of 97.79%. Additionally, it achieved a mean AUC-PR value of 0.991 and an average AUC-ROC score of 0.997 [12].

As a pre-processing step, the OMLTS-DLCN model used an adaptive fuzzy based median filtering (AFF) technique with the Shell Game Optimization (SGO) algorithm, which had a higher accuracy of 98.50% and 97.55% on the Mini-MIAS dataset and DDSM dataset, respectively [13]. A CNN model called CoroNet was suggested for use in the automatic detection of Breast Cancer using the CBIS-DDSM dataset. It used the Xception architecture, which had an overall accuracy of 94.92% after being pre-trained on the Image-Net [14]. Using a dataset provided by UCI and an F1 score of greater than 98, a study was conducted with the goal of developing a deep neural network that could predict the malignancy of breast cancer [15]. The categorization capacity of breast tumors applied to ultrasound pictures was seen in a study taking into account eight different fine-tuned pre-trained models. With regard to several performance measures, they developed a shallow bespoke convolutional neural network that performs better than the pre-trained models. In contrast to the best pre-trained model, which displays 92% accuracy and 0.972 AUC score, they estimated to attain 100% accuracy and achieve 1.0 AUC score [16].

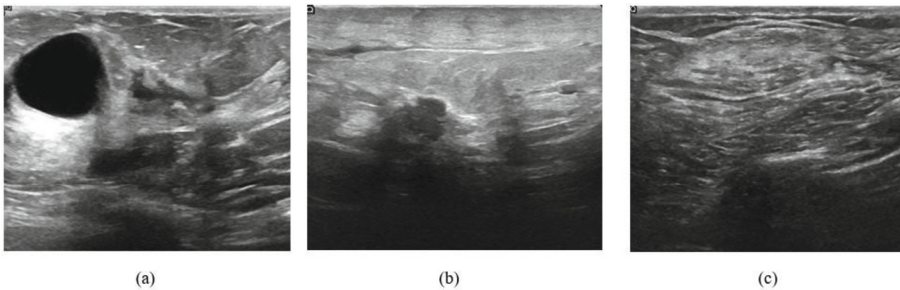
A study using the segmentation model ResU-segNet and the hierarchical fuzzy classifier HFC, which combined the interval type-2 probabilistic fuzzy c-means (IT2PFCM) and fuzzy neural network (FNN), revealed a relatively low accuracy of 91% [17]. A development of two techniques for visualizing important nuclei in classification of histology

images using graph convolutional networks had been shown with a range of accuracies between 92% and 94% [18]. An efficient classification model with a high performance up to 99.05% accuracy value had been researched using hyper-parameter optimization with parallel computing architecture and CUDA-enabled graphics processing unit [19]. An implementation of pooling structures had been applied on most CNN-based models, which might greatly improve the models' performance on mammographic image data with the same input, showed their overall accuracy close to 85% [20]. An implementation of deep learning algorithms using a conventional CNN and a recurrent CNN to differentiate three breast cancer molecular subtypes on MRI, had been shown where mean accuracy was close to 91% [21]. Again there were quite similar kind of implementations of CNNs, Res-Net50 and other neural networks related DL models by which they range of accuracies close to 85% to 90% [22–26].

Accordingly, we proposed a new convolutional neural network model as CEIMVEN with the implementation of modified Efficient-Net-V1 (b0-b7) and Efficient-Net-V2 (b0-b3). Our method came across with a far better overall accuracy than most other previous studies could possibly pull off. The descriptive representations about the accuracy and total result analysis can be found at the section of "Results and Accuracies".

### 3 Dataset and Proposed Method

For the extensive amount of data processing, We had to search through the database of live, open-source ultrasound images, collected from 600 female patients in 2018, aged between 25 and 75 years old [27]. Exactly 780 photos with an average size of  $500 \times 500$  pixels made up the collection. Nevertheless, we executed a full-on data augmentation procedure for having almost 4500 more images from it for better training accuracies. All of the images were formatted into PNG whereas the ground truth images are presented with original images. Those images were categorized into 3 classes, which were identified as benign, malignant and normal; demonstrated in Fig. 1.



**Fig. 1.** Classified Ultrasound Image Dataset with 3 classes, identifying as stages of cancer; (a) benign (b) malignant and (c) normal

Our proposed CEIMVEN method is detailed within two main components and they are: training the breast cancer detection and classification model with modified (a) EfficientNet-V1 with all of its sub-components (b0, b1, b2, b3, b4, b5, b6 and b7)

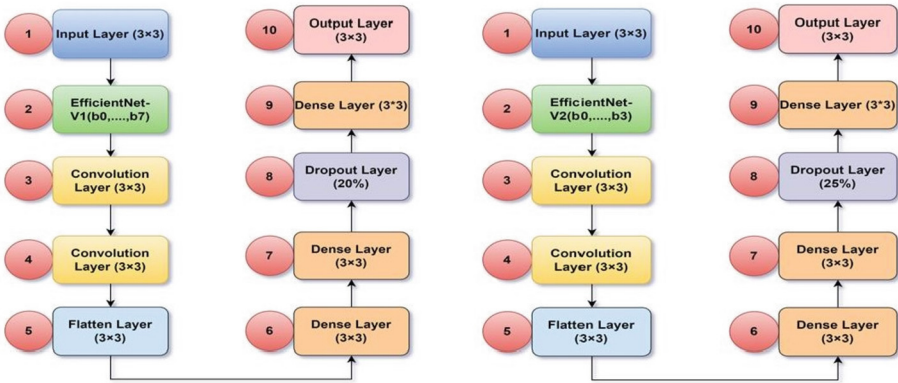
and (b) the cutting-edge EfficientNet-V2 with all of its latest sub-components (b0, b1, b2 and b3).

At first, we had to fulfill and settle down with our complex requirement of computer vision related software and hardware machineries. For the deep learning and image processing environment we chose to utilize Pycharm IDE and train the deep learning (DL) model locally. To make our model compatible with using the same model on online platforms in the future, we saved our model as an H5 file and translated it to TensorFlow-JavaScript. We utilized CVAT for image data labeling with bounding boxes with the mask value we got from the open-source dataset for identifying the specific location of cancer cells from the ultrasound image. We were using the masking values as our reference point. Version 4.5.3.56 of OpenCV-Python and Python interpreter 3.9.1 with the majority of its global packages and variables were used. As a model backend, Keras in version 2.10.0 was employed. The main architectures of EfficientNet-V1 (b0-b7) and EfficientNet-V2 (b0-b3) were imported using Tensorflow version 2.6.1. We used an Intel Core i5- 10th Generation CPU with a clock speed of 2.11 GHz to be more precise with our gear. For the NVIDIA Geforce MX110 with 2GB Graphics Card, we installed CUDA version 10.2 and CUDA deep neural network libraries (CuDNN) for its largely compatible GPU support.

In Fig. 2, we manifested the implementation of our own neural network architecture with more improvisation and modification within conventional EfficientNet-V1 with 10 different layers. At first, we initiated an input layer of  $3 \times 3$  filter size. Then we called the core pre-trained model of EfficientNet-V1. All of the sub-components such as: b0, b1, b2, b3, b4, b5, b6 and b7 were added and implemented gradually and iteratively with TensorFlow and Keras. We approached transfer learning methodology by using the pre-trained model as its core, fine-tuning it with more hyper-parameters and increasing its training and validation accuracies for a better research outcome.

After calling the core model, the input images moved from one convolution (Conv2D) layer with a  $3 \times 3$  pixel entity filter size to another convolution (Conv2D) layer with a  $3 \times 3$  pixel entity filter size. Conv2D's hidden layers started off with a filter size of  $3 \times 3$  for the current preprocessing. Conv2D filter numbers were modified to 256 and 128 filters, respectively, in separate layers. In these parameters, the strides, padding, and activation functions were left unchanged. For effectively triggering multiple classification, we summoned the Rectified Linear Unit (ReLU) function. Then, using Conv2D, Dropout and Flatten layers as well as GlobalAveragePooling2D functions, the hyper-parameters were effectively optimized. For three distinct forms of stage detection (Benign, Malignant, and Normal), three (3) classes were established. We preprocessed and downsized the photos to a  $224 \times 224$  pixel impression. Through the Flatten layer, the two dimensional tensor array values were transformed into one dimensional picture data. We added a Dropout layer as the 8<sup>th</sup> layer of the architecture. We had to initiate a dropout of 20% for keeping away the under-fitting scenarios of our DL model. We added another Dense layer having filter size of  $3 \times 3$  for getting the two-dimensional array values into a single dimension.

Lastly, we added an output layer with filter size of  $3 \times 3$ . 'Softmax' function was utilized for multi-class output. We could have used 'Sigmoid' function if we needed a binary classification only (Benign and Malignant). But as we were approaching for 3

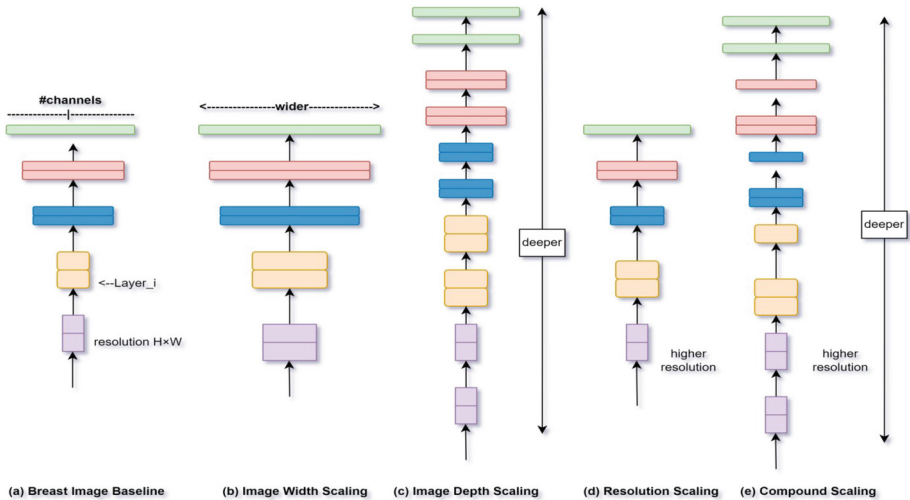


**Fig. 2.** The architecture of our proposed modified EfficientNet-V1 (b0-b7) and EfficientNet-V2 (b0-b3) model for cancer classification with 5% increase at V2 dropout (10 layers at each)

classes, we used 'Softmax'. After we processed image data augmentation, we had to deal to a larger amount of image data. A significant number of epoch were also required for parameter adjustment, training, and testing on the large size of picture data. Device-installed GPU handled larger model training more quickly, although TensorFlow-GPU was also deployed as a backup. The batch size was 42, the epoch number was 500, and the initial learning rate was 0.01. To define vanishing gradient properties, stochastic gradient descent (SGD) was turned on. We used 4,100 picture samples for training, 200 for validation, and 200 for testing. We trained and saved the improvised EfficientNet-V1 (b0-b7) models as .h5 files and .pkl artifacts for further training and testing.

Quite similarly in the right part of Fig. 2, we demonstrated another implementation of our own neural network architecture with more improvisation and modification within conventional EfficientNet-V2 with another 10 different layers. At first, we initiated an input layer with filter size of  $3 \times 3$ . Then we called the core pre-trained model of EfficientNet-V2. All of the latest and cutting-edge sub-components such as: b0, b1, b2 and b3 were added and implemented gradually and iteratively with TensorFlow and Keras. We approached transfer learning methodology by using the pre-trained model as its core, fine-tuning it with more hyper-parameters and increasing its training and validation accuracies. After calling the core model, the input images moved from one convolution (Conv2D) layer with  $3 \times 3$  pixel entity filter size to another convolution (Conv2D) layer with a  $3 \times 3$  pixel entity filter size. Conv2D's hidden layers started off with filter size of  $3 \times 3$  for the current preprocessing. Conv2D filter numbers were modified to 256 and 128 filters, respectively, in separate layers. In these parameters, the strides, padding, and activation functions were left unchanged. For effectively triggering multiple classification, the Rectified Linear Unit (ReLU) function was called. Then, using Conv2D, Dropout and Flatten layers as well as GlobalAveragePooling2D functions, the hyper-parameters were effectively optimized. For three distinct forms of stage detection (Benign, Malignant, and Normal), three (3) classes were established. We preprocessed and downsized the photos to a  $224 \times 224$  pixel impression. Through the Flatten layer, the 2D tensor values were transformed into 1D picture data. We added a Dropout layer

as the 8<sup>th</sup> layer of the architecture. We had to initiate a dropout of 25% for keeping away the under-fitting scenarios of our DL model which was significantly 5% higher than the previous EfficientNet-V1. We added another Dense layer with filter size of  $3 \times 3$  for getting the two dimensional array values into a single dimension. Lastly, we added an output layer with filter size of  $3 \times 3$ . ‘Softmax’ function was utilized for multi-class output. After we processed image data augmentation, we had to deal to a larger amount of image data. A significant number of epoch were also required for parameter adjustment, training, and testing on the large size of picture data. Device-installed GPU handled larger model training more quickly, although TensorFlow-GPU was also deployed as a backup. The batch size was 42, the epoch number was 500, and the initial learning rate was 0.01. To define vanishing gradient properties, stochastic gradient descent (SGD) was turned on. We used 4,100 picture samples for training, 200 for validation, and 200 for testing. We trained and saved the improvised EfficientNet-V1 (b0-b7) models as .h5 files and .pkl artifacts for further training and testing.



**Fig. 3.** EfficientNet Model scaling for our proposed method

From Fig. 3 we can observe that (a) is a baseline network example for the breast ultrasound images. From (b) to (d) are conventional scaling that only increases one dimension of our neural network's width, depth and resolution. Therefore, (e) is our proposed compound neural scaling method which executes the uniform scaling of all three dimensions within a fixed aspect ratio. This is exactly what the EfficientNet-V1-V2 model stands for, showing its unique amalgamation of compound scaling in different scaling aspects and scenarios.



## 4 Results and Accuracies

As we entered into our CEIMVEN model testing phase, we divided our tasks into several steps such as: (a) recording the Area Under ROC Curve (AUC) of training and validation, (b) recording loss values of training and validation and (c) taking steps to improve the result values coming from the DL models.

**Table 1.** Modified EfficientNet-V1 models’ average AUC (training, validation) for proposed breast cancer classification

Model	Training Accuracy (Approx.)	Validation Accuracy (Approx.)
EfficientNet-V1-b0	99.45%	82.47%
EfficientNet-V1-b1	99.21%	82.73%
EfficientNet-V1-b2	99.29%	83.89%
EfficientNet-V1-b3	99.67%	82.93%
EfficientNet-V1-b4	99.85%	84.31%
EfficientNet-V1-b5	99.79%	84.86%
EfficientNet-V1-b6	99.83%	85.27%
EfficientNet-V1-b7	99.89%	85.68%

In Table 1, we represented the average AUC in training and validation while we executed the DL model training with EfficientNet-V1 and its sub-components. Here, we can see the training accuracies of EfficientNet-V1’s b0 to b7 fluctuates between different decimal values of 99%. Notably the training accuracy rose very negligibly higher from b0 to b7. While validation accuracies fluctuates between the values of 82.47% to 85.68%, signifies a good sign of accuracies than other deep neural nets like Inception-Net-V2 or ResNet-V1-V2.

In Table 2, we represented the average loss in training and validation while we executed the DL model training with EfficientNet-V1 and its sub-components. Here, we can see the training losses of EfficientNet-V1’s b0 to b7 fluctuates between the values of 0.022 to 0.035. Notably the training losses fluctuated quite inconsistently between b0 and b7. While validation losses fluctuated between the values of 0.511 to 0.594, signifies quite a good gap of losses. Simultaneously the losses shows significantly lower in numbers than other deep neural nets like Inception-Net-V2, VGG16-19 or ResNet-V1-V2.

In Table 3, we represented the average AUC in training and validation while we executed the DL model training with EfficientNet-V2 and its sub-components. Here, we can see the training accuracies of EfficientNet-V2’s b0 to b3 fluctuates between different decimal values of 99%, almost same as its predecessor V1. Notably the training accuracy rose very negligibly higher from b0 to b3. While validation accuracies fluctuates between the values of 82.47% to 85.43%, signifies a good sign of accuracies than other deep neural nets like VGG16-19, Inception-Net-V2 or ResNet-V1-V2.

In Table 4, we represented the average loss in training and validation while we executed the DL model training with EfficientNet-V2 and its sub-components. Here, we

**Table 2.** Modified EfficientNet-V1 models' average Loss (training, validation) for proposed breast cancer classification

Model	Training Loss (Approx.)	Validation Loss (Approx.)
EfficientNet-V1-b0	0.035	0.582
EfficientNet-V1-b1	0.022	0.590
EfficientNet-V1-b2	0.028	0.511
EfficientNet-V1-b3	0.032	0.573
EfficientNet-V1-b4	0.029	0.582
EfficientNet-V1-b5	0.033	0.590
EfficientNet-V1-b6	0.028	0.586
EfficientNet-V1-b7	0.030	0.594

**Table 3.** Modified EfficientNet-V2 models' average AUC (training, validation) for proposed breast cancer classification

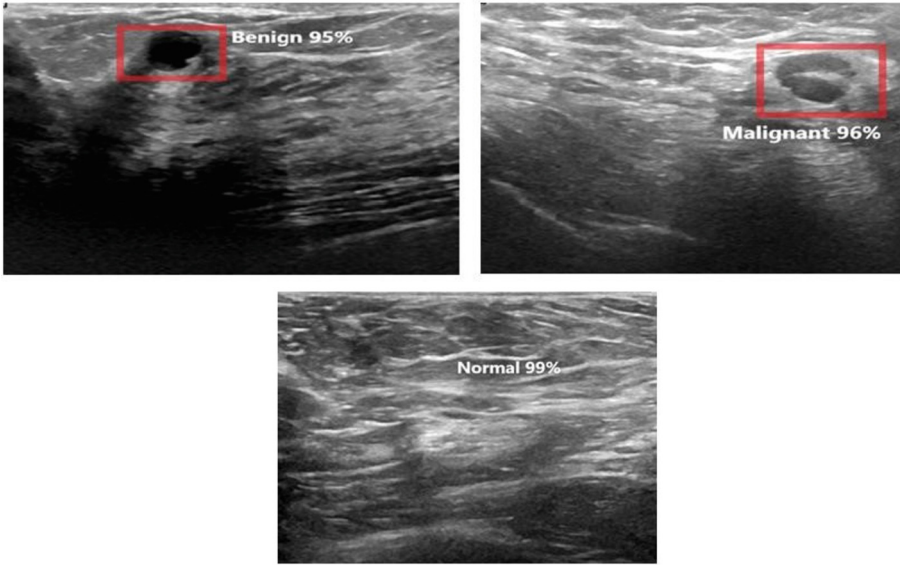
Model	Training Accuracy (Approx.)	Validation Accuracy (Approx.)
EfficientNet-V2-b0	99.45%	82.47%
EfficientNet-V2-b1	99.21%	83.73%
EfficientNet-V2-b2	99.49%	84.89%
EfficientNet-V2-b3	99.87%	85.43%

**Table 4.** Modified EfficientNet-V2 models' average loss (training, validation) for proposed breast cancer classification

Model	Training loss (Approx.)	Validation loss (Approx.)
EfficientNet-V2-b0	0.028	0.582
EfficientNet-V2-b1	0.032	0.590
EfficientNet-V2-b2	0.029	0.586
EfficientNet-V2-b3	0.033	0.594

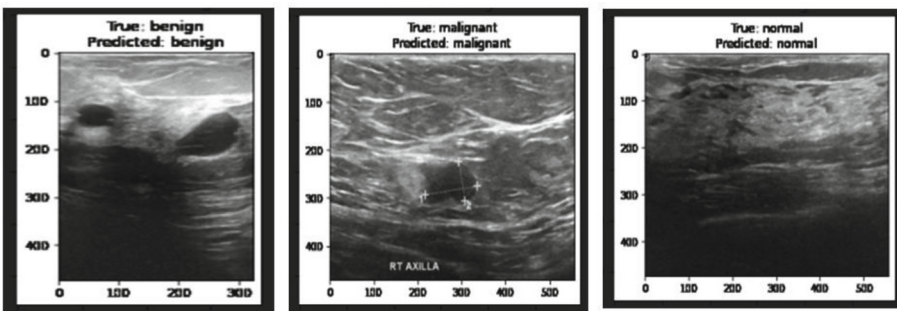
can see the training losses of EfficientNet-V2's b0 to b3 fluctuates between the values of 0.028 to 0.033, almost same as its predecessor V1. Notably the training losses fluctuated quite less inconsistently between b0 and b3. While validation losses fluctuated between the values of 0.582 to 0.594, signifies quite a less gap of losses. Simultaneously the losses shows significantly lower in numbers than other deep neural nets like Mobile-Net, Inception-Net-V2 or ResNet-V1-V2.

As we labelled the ultrasound breast images with CVAT upon its masked reference values, it helped us training another channel of the DL model as an object detector. Here,



**Fig. 4.** Breast cancer cell detection from ultrasound image as “Benign” with 95% confidence, “Malignant” with 96% confidence and “Normal” with 99% confidence.

we used the cancer affected areas as the object which we wanted to detect and detect the class of the cancer simultaneously. We used the same image resolution for training and testing with a shape of  $224 \times 224$  pixels. We tried to maintain the input layer of same  $3 \times 3$  filter size and a single output layer with filter size of  $3 \times 3$ . We executed the input of the images one-by-one and got the prediction results with rectangular bounding boxes, identifying the cancer affected areas. The prediction results have been shown in Fig. 4.



**Fig. 5.** Testing the prediction of the classes (benign, malignant and normal) from ultrasound images without identifying any cancer cell.

In Fig. 5, we demonstrated the classification and prediction capability of our implemented breast cancer DL model with the comparison from “true” to “predicted” landmark with a 2D graphical visualization. Here, we emphasized more on the cancer classification

part rather than the detection of the cancer cell and affected breast areas. Combining all of the test cases the average prediction accuracy was approximately 97.79%. We dealt with a generative confusion matrix where concerned about the four important factors like: True Positive, True Negative, False Positive and False Negative. The average precision rate was approximately 1.36% while the recall rate was approximately 0.84%. All of these detection and prediction result factors for each classes are demonstrated in Table 5.

**Table 5.** Accuracies of Prediction and Detection, Precision and Recall rate for each classes:

Model	Prediction Accuracy (Approx.)	Precision Rate (Approx.)	Recall Rate (Approx.)
Benign	97.86%	1.89%	0.25%
Malignant	96.27%	1.95%	1.78%
Normal	99.26%	0.25%	0.49%

At first, the average training accuracy for affected area detection was over 97%, while the average testing accuracy was over 91.47%. With such parameters as "re-scale" = 1/255, "horizontal-vertical flip", "rotation range" = 360, "width-height" shift, and "zoom-brightness" range with its important aperture, we attempted to recreate a number of picture data augmentation functions. We managed to renew the image data and materially improve the dataset. We used the same iteration approach and epoch numbers to retrain the model. The retesting procedure was started by us. The testing accuracy thereafter had a huge increase, ending at 96.49% on an average landmark. We noticed much greater accuracy when we increased the batch size to 54 and the epoch number to 600. Overall, the CNN-algorithmic improvised EfficientNet-V1-V2 based ultrasound image data model training and prediction testing performed with an average accuracy rate of 97.79% allowed us to carry out the integrated process of detection and prediction.

## 5 Conclusion and Future Endeavors

Our main focus in this study was the real-time identification and classification of breast cancer from ultrasound pictures using more recent and state-of-the-art deep neural network designs. We wanted to break the stereotype of using mammograms from the medical image analysis and expecting better accuracies from that. We wanted to embed state of the art neural models into so called obsolete ultrasound images and drag the best prediction accuracies out of it. Undoubtedly we pulled off a very bold and challenging research work and significantly rich results out of it. No previous research work has ever tried the modification of EfficientNet-V1 (b0-b7) and EfficientNet-V2 (b0-b3) with a custom DL model along ultrasound images. That's why we approached this method for the creation of a unique research austerity and achievement of a significantly higher overall accuracy (97.79%). With overall data augmentation we used a total number of

5280 images while we partitioned the entire dataset into (train, validation, test) portions as (80%, 10%, 10%) channels. 10% of the testing data were precisely used one-by-one as input for getting the prediction results.

In the future, we will strive to collaborate with more robust datasets from various reputed cancer institutes in the world.

We are confident that adding more patient image datasets would improve the capabilities and prediction accuracy of our DL artifacts. In order to improve the diagnosis of breast cancer with higher accuracies while taking into account cutting-edge techniques like Inception-ResNet-V2, Fast-AI, Dense-Net (121, 169, 201) and so on, we have significant plans to expand our vigorous research work.


## References

1. Dammu, H., Ren, T., Duong, T.Q.: Deep learning prediction of pathological complete response, residual cancer burden, and progression-free survival in breast cancer patients. *Plos one* **18**(1), e0280148 (2023)
2. Debien, V., et al.: Immunotherapy in breast cancer: an overview of current strategies and perspectives. *NPJ Breast Cancer* **9**(1), 7 (2023)
3. Swain, S.M., Shastry, M., Hamilton, E.: Targeting HER2-positive breast cancer: Advances and future directions. *Nature Reviews Drug Discovery* **22**(2), 101–126 (2023)
4. Dongsar, T.T., Dongsar, T.S., Abourehab, M.A., Gupta, N., Kesharwani, P.: Emerging application of magnetic nanoparticles for breast cancer therapy. *European Polymer Journal*, 111898 (2023)
5. Sánchez-Cauce, R., Pérez-Martín, J., Luque, M.: Multi-input convolutional neural network for breast cancer detection using thermal images and clinical data. *Computer Methods and Programs in Biomedicine*, **204**, 106045 (2021)
6. Chen, P.Y., et al.: Automatic breast tumor screening of mammographic images with optimal convolutional neural network, *Applied Sciences* **12**(8), 4079 (2022)
7. Oyelade, O.N., Ezugwu, A.E.: A novel wavelet decomposition and transformation convolutional neural network with data augmentation for breast cancer detection using digital mammogram, *Scientific Reports* **12**(1), 5913 (2022)
8. Khan, S.I., Shahrir, A., Karim, R., Hasan, M., Rahman, A.: MultiNet: a deep neural network approach for detecting breast cancer through multi-scale feature fusion, *Journal of King Saud University-Computer and Information Sciences* **34**(8), 6217–6228 (2022)
9. Altameem, A., Mahanty, C., Poonia, R.C., Saudagar, A.K.J., Kumar, R.: Breast cancer detection in mammography images using deep convolutional neural networks and fuzzy ensemble modeling techniques, *Diagnostics* **12**(8), 1812 (2022)
10. Aljuaid, H., Alturki, N., Alsubaie, N., Cavallaro, L., Liotta, A.: Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning, *Computer Methods and Programs in Biomedicine* **223**, 106951 (2022)
11. Sureka, M., Patil, A., Anand, D., Sethi, A.: Visualization for histopathology images using graph convolutional neural networks. In: 2020 IEEE 20th international conference on bioinformatics and bioengineering (BIBE), pp. 331–335. IEEE (2020)
12. Hameed, Z., Garcia-Zapirain, B., Aguirre, J.J., Isaza-Ruget, M.A.: Multiclass classification of breast cancer histopathology images using multilevel features of deep convolutional neural network, *Scientific Reports* **12**(1), 15600 (2022)
13. Kavitha, T., et al.: Deep learning based capsule neural network model for breast cancer diagnosis using mammogram images, *Interdisciplinary Sciences: Computational Life Sciences*, 1–17 (2021)

14. Mobark, N., Hamad, S., Rida, S.Z.: Coronet: Deep neural network-based end-to-end training for breast cancer diagnosis, *Applied Sciences* **12**(14), 7080 (2022)
15. Qi, X., Hu, J., Zhang, L., Bai, S., Yi, Z.: Automated segmentation of the clinical target volume in the planning CT for breast cancer using deep neural networks, *IEEE Transactions on Cybernetics* **52**(5), 3446–3456 (2020)
16. Masud, M., Eldin Rashed, A.E., Hossain, M.S.: Convolutional neural network-based models for diagnosis of breast cancer, *Neural Computing and Applications*, 1–12 (2020)
17. Shen, T., Wang, J., Gou, C., Wang, F.Y.: Hierarchical fused model with deep learning and type-2 fuzzy learning for breast cancer diagnosis, *IEEE Transactions on Fuzzy Systems* **28**(12), 3204–3218 (2020)
18. Shu, X., Zhang, L., Wang, Z., Lv, Q., Yi, Z.: Deep neural networks with region-based pooling structures for mammographic image classification, *IEEE transactions on medical imaging* **39**(6), 2246–2255 (2020)
19. Burçak, K.C., Baykan, Ö.K., Uğuz, H.: A new deep convolutional neural network model for classifying breast cancer histopathological images and the hyperparameter optimisation of the proposed model, *The Journal of Supercomputing* **77**, 973–989 (2021)
20. Prakash, S.S., Visakha, K.: Breast cancer malignancy prediction using deep learning neural networks. In: 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 88–92. IEEE (2020)
21. Zhang, Y., et al.: Prediction of breast cancer molecular subtypes on DCE-MRI using convolutional neural network with transfer learning between two centers, *European radiology* **31**, 2559–2567 (2021)
22. Desai, M., Shah, M.: An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN), *Clinical eHealth* **4**, 1–11 (2021)
23. Al-Haija, Q.A., Adebajo, A.: Breast cancer diagnosis in histopathological images using ResNet-50 convolutional neural network. In: 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), pp. 1–7. IEEE (2020)
24. Ahmed, L., et al.: Images data practices for semantic segmentation of breast cancer using deep neural network, *Journal of Ambient Intelligence and Humanized Computing*, 1–17 (2020)
25. Lan, K., et al.: Multi-view convolutional neural network with leader and long-tail particle swarm optimizer for enhancing heart disease and breast cancer detection, *Neural Computing and Applications* **32**, 15469–15488 (2020)
26. Zheng, J., et al.: Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis, *IEEE Access* **8**, 96946–96954 (2020)
27. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. *Data Brief* **28**, 104863 (2020). <https://doi.org/10.1016/j.dib.2019.104863>



# Performance Improvement of Movie Recommender System Using Spectral Bi-clustering with Mahalabonis Distance

Sonu Airen<sup>(✉)</sup>  and Jitendra Agrawal

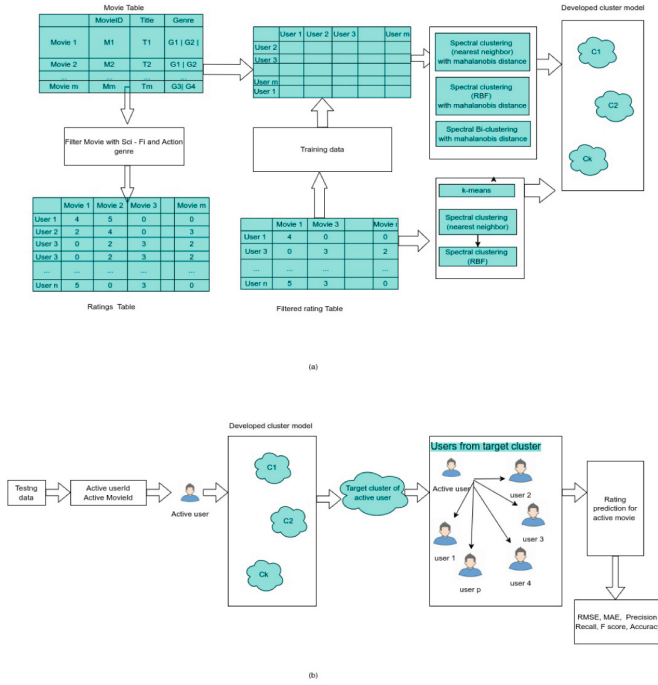
SOIT, RGPV, Bhopal, M.P., India  
sonugoyal24@rediffmail.com

**Abstract.** Collaborative Filtering with clustering is the primary method of recommendation. This research work introduced a new Spectral Bi-clustering with the Mahalabonis distance-based Movie Recommendation algorithm for Collaborative Filtering. In this paper, we compare the performance of several clustering methods like Kmeans, Spectral clustering with radial Basis Function and nearest neighbors affinity, Spectral clustering with radial Basis Function and nearest neighbors affinity with Mahalabonis distance, Spectral Bi-clustering with Mahalabonis distance to generate a movie recommendation system. Our experimental results show a significant performance improvement of our proposed Spectral Bi-clustering with the Mahalanobis distance-based Movie Recommendation algorithm over the traditional K-means algorithm as well as the Spectral Clustering algorithm with nearest neighbors and radial Basis Function affinity for Movie Recommender System.

**Keywords:** Accuracy · Mahalabonis distance · Precision · Recall · Recommender systems · Spectral clustering · Spectral Bi-clustering

## 1 Introduction

Synthesizing clustering with Collaborative Filtering [2, 11] is a model-based Recommender System technique that is used to reduce computation time. The main task of the Collaborative Filtering algorithm is to calculate the similarity among items or users. For finding similarities same tastes users are grouped in clusters. The clustering phase is performed offline so the computation efficiency of the system increases. After applying clustering the neighborhood of active users is found by searching the most suitable cluster for active users. The members of the cluster are now the neighborhood of active users and are known as sim-users of active users. The predicted rating for active users is calculated from the average ratings of its neighborhood users who co-rated the active item. In this paper, we compare performance of several clustering methods like Kmeans, Spectral clustering with nearest neighbors (SC-NN), Spectral clustering with nearest neighbors affinity with Mahalanobis distance (SC-NN-MD), Spectral clustering with



**Fig. 1.** Movie recommender system using clustering (a) Clustering model generation. (b) Recommendation generation.

radial Basis Function affinity with Mahalanobis distance (SC-RBF-MD), Spectral Bi-clustering with Mahalanobis distance (SC-BI-MD) to generate a movie recommendation system.

The rest of the paper is organized as follows: Existing works related to different clustering algorithms, Mahalanobis distance is discussed in Sect. 2. In Sect. 3, the proposed approach for Movie Recommender System using Spectral Bi-clustering with Mahalanobis distance is discussed in detail. In Sect. 4, preliminaries are discussed. In Sect. 5, Dataset, evaluation metrics and performance observation of the proposed system about various clustering algorithms on a standard dataset are reported. Lastly, concluding remarks are presented in Sect. 6.

## 2 Related Work

Movie Recommender System-related existing work based on different clustering algorithms like K-means, Spectral clustering, Spectral bi-clustering and Mahalanobis distance are discussed in this section.

The concept of mean rating, normalized rating, and baseline rating with simple K Nearest Neighbors (KNN) algorithm have been used in [3], to describe different varieties of the KNN-based CF algorithm. The aforementioned techniques



were further tested for neighborhood construction using four distinct similarity measures: Mean Square Difference, Cosine, Pearson, and Pearson Baseline Similarity. In [4], authors used the K-means algorithm for movie recommendation. The author proposed the RecTree method for clustering users as well as items and achieved numerosity reduction and feature selection. In clus.Sarwar, the author proposed clustering techniques in collaborative filtering recommendations to improve the scalability of recommender systems. Authors claimed that clustering may reduce the accuracy of the recommendation system but increase the throughput of the system. In [2], the authors developed a K-Means Clustering and K-Nearest Neighbour algorithm-based movie recommender system and demonstrated that the RMSE value of the suggested system achieves the same value as the current technique while utilizing fewer clusters.

In [6] author shows that theoretically speaking, a weighted graph partitioning objective is identical to a general weighted kernel k-means objective. The authors demonstrate that in situations where eigenvector computing is impractical, weighted kernel k-means is used for graph partitioning. Additionally, the author demonstrates that non-spectral methods of graph splitting are just as successful as spectral methods. In [12] the authors explain various spectral clustering techniques, graph Laplacians, and their fundamental features. They also give methods for constructing these algorithms from scratch. In [10] an easy spectral clustering approach is presented by the author. The matrix perturbation theory is used to analyze the algorithm, and favorable conditions are provided. In several difficult clustering situations, experimental results are presented.

The author in [8], created the Spectral Bi-clustering approach, which concurrently groups genes and diseases while identifying a different checkerboard structure in a matrix of gene expression data. They applied Spectral Bi-clustering to the cancer expression data set and examined the checkerboard structure to find the particular type of tumors in cancer patients. ScuBA clustering is proposed by [1], which is based on subspace clustering to handle the sparsity and high dimensionality problems of input data. In [14] authors proposed a Multi-class Co-Clustering problem by creating several subgroups by Co-Clustering users and items to provide top-N Recommendations. In [13], authors handled the sparsity problem of recommendation by smoothing the missing values of the rating matrix with the help of Bi-clustering. The authors propose Levelwise clustering [7] to handle the sparsity problem. Users and products are clustered separately using pairwise similarity in the first level of clustering. At the second level bipartite graph, Spectral Co-clustering of user-item is performed. Several conventional approaches to bi-clustering are analyzed by the authors [9], who then categorize them based on the kinds of bi-clusters they can identify, the patterns of bi-clusters they've found, the techniques they used to search, and the applications they hope to utilize them for.

### 3 Proposed System Architecture

The proposed system architecture for Movie Recommender System Using Spectral Bi-Clustering with Mahalanobis distance is shown in Fig. 1. Recommendations

are generated using historical user rating data. The paper's suggested recommendation algorithm utilizes a cluster-only methodology. Creating a data model i.e. offline user cluster development, and online rating prediction generation, are the two parts that make up this process. In the first section, clustering techniques are utilized, and the following recommendations are given solely based on the built model. In this paper, Spectral Bi-clustering with Mahalanobis distance is proposed for clustering. For this purpose, movies with two genres, namely action and sci-fi, are considered, and users are clustered using these two genres' average ratings only. For SC-RBF-MD, SC-NN-MD and SC-BI-MD pairwise Mahalanobis distance is calculated Among users. Preprocessed data is clustered using K-means, SC-NN, SC-RBF-MD, SC-NN-MD and SC-BI-MD for different values of clusters, i.e.  $k = 2$  to 40. The user and Movie for whom the recommendation is required is known as an active users and active movies in this paper. For rating prediction active user cluster is searched and members of the cluster become active user neighbors. The average rating from neighbors becomes the predicted rating for the active user. Metrics like RMSE, MAE, Precision, Recall, accuracy and F-score are plotted for different numbers of clusters.

## 4 Preliminaries

A common unsupervised learning technique is clustering. A clustering is a grouping of data points, in mutually exclusive subsets so that points in a cluster have a greater degree of similarity to one another than to points in other clusters. Suppose  $D = \{D_1, D_2, \dots, D_n\}$  is set of data points.  $D_i = \{D_{i1}, D_{i2}, \dots, D_{im}\}$  is  $m$ -dimensional  $i^{th}$  data point. Clustering forms  $k$  clusters from these data points i.e.  $C = \{C_1, C_2, \dots, C_K\}$  in such a manner that  $C_k \cap C_l = \phi$  and  $\cup_{k=1}^K C_k = D$ .

### 4.1 Mahalanobis Distance

Mahalanobis distance [5] takes into account the correlation between variables.

$$D^2 = (x - m)^T \cdot C^{-1} \cdot (x - m) \quad (1)$$

where  $x$  = vector of data

$D^2$  = Mahalanobis distance

T = indicates vector should be transported

$C^{-1}$  Inverse Covariance matrix of independent variables

$m$  = Vector of mean values of independent variables

### 4.2 Bi-clustering

Bi-clustering techniques [6,8] are simultaneously clustering on the rows and columns dimension of the rating matrix. This simultaneous clustering seeks to find sub-matrices that are a subgroup of users and the subgroup of items, where the user exhibits highly co-rated ratings for every item. Bi-clustering techniques

can be used in collaborative filtering to find the subset of users having comparable behavior towards a subset of items. By grouping users with similar tastes on similar items we can cluster them into the same cluster and predict their interest and make proper predictions. suppose  $R = (P, Q)$  and  $P = \{p_1, p_2, \dots, p_m\}$  and  $Q = \{q_1, q_2, \dots, q_n\}$ . If  $I \subseteq P$  and  $J \subseteq Q$  i.e.  $I$  is a subset of rows and  $J$  is a subset of columns.  $R_{I,J} = (I, J)$  denotes the submatrix  $R$  that contains only the elements  $r_{i,j}$  belongs to the submatrix with a set of rows  $I$  and a set of columns  $J$ . A bichuster  $(I, J)$  can be defined as a  $k$  by  $s$  submatrix of the rating matrix  $R$  where  $(k \leq m)$  and  $(s \leq n)$ .

### 4.3 Spectral Clustering

The data points are considered as nodes of a network in spectral clustering [6, 10, 12]. The treatment of clustering as a graph partitioning problem follows. The nodes are represented here in a low-dimensional space that is easily separable into clusters using eigenvalues and eigen vectors. Data point  $x_1, x_2, \dots, x_n$  is represented as a vertex of similarity graph  $G = (V, E)$ .  $s_{ij}$  is an edge between the data points  $x_i$  and  $x_j$  if similarity  $s_{ij}$  is positive or larger than a certain threshold.

---

#### Algorithm 1: Spectral clustering

---

**Input:**

Graph  $G$  with  $V$  vertices and  $E$  edges.  
 $S$  is similarity matrix of  $G$ .

**Output:**

$C = \{C_1, C_2, \dots, C_k\}$  a set of clusters.

**Begin:**

**calculate** affinity matrix using Eq. 2  
**calculate** degree matrix( $D$ ) of  $G$  using Eq. 3  
**calculate** Laplacian matrix ( $L$ ) of graph  $G$  using Eq. 4  
**find**  $k$  largest eigenvectors  $\{u_1, u_2, \dots, u_k\}$  of  $L$  and  
**form** the matrix  $P$  using eigenvector  $\{u_1, u_2, \dots, u_k\}$   
as columns  $\in n \times k$   
**represent** every node by the corresponding row of new matrix  $P$ .  
**cluster**  $X$  into  $k$  clusters via K-means  
**assign** the original point  $x_i$  to cluster  $j$   
iff row  $i$  of matrix  $P$  was assigned to cluster  $j$

**End.**

---

Spectral clustering is based on the spectrum of an affinity matrix. Affinity is directly proportional to similarity. Affinity matrix weighted edges  $A_{ij}$  are computed by some local symmetric and non-negative similarity measures like  $\epsilon$ -neighborhood graph approach,  $k$ -nearest neighbors approach and Gaussian kernel. In the  $\epsilon$ -neighborhood graph approach  $\epsilon$  is identified and for affinity value greater than  $\epsilon$ , the edge is included between data points. Edges are inserted

between a node and its k-nearest neighbors in the k-nearest neighbors approach. Gaussian kernel forms a fully connected graph by inserting an edge between every pair of nodes by calculating similarity using Eq. 2:

$$A_{ij} = \exp - \frac{\|v_i - v_j\|^2}{2\sigma^2} \tag{2}$$

if  $i \neq j$  and  $A_{ii} = 0$  where  $\sigma^2$  is scaling parameter.

D is  $n \times n$  degree matrix of G. Degree of vertex  $v_i$  is the total weight of edges incident to vertex  $v_i$

$$D(i, i) = \sum_j v_{ij} \tag{3}$$

L is  $n \times n$  symmetric Laplacian matrix (L) of graph G is

$$L = D - A \tag{4}$$

Graph partitioning objective is a function of the edge cut of the partition. Several different graph partitioning objectives have been proposed and studied such as:

Min-Cut: Partition graph into sets such that the weight of edges connecting vertices in one partition to vertices in another partition is minimum.

$$cut(V1, V2) = \sum_{i \in V1} \sum_{j \in V2} w_{ij} \tag{5}$$

Ratio Cut: It seeks to minimize the cut between clusters and the remaining vertices. It uses the size of the partition and is expressed as

$$Ratio\_Cut(G) = \sum_{c=1}^k \frac{cut(V, V \setminus V_c)}{|V_c|} \tag{6}$$

Normalized Cut: Spectral clustering used a relaxed form of normalized cut of the graph. Normalized Cut tries to balance the size of the partition by using the volume(sum of the degree of vertices in the partition) of the partition instead of size to form an equal-size partition.

$$Norm\_Cut(G) = \sum_{c=1}^k \frac{cut(V, V \setminus V_c)}{degree(V_c)} \tag{7}$$

A relaxed form of normalized cut of graph laplacian is used to solve the NP-hard problem of graph partitioning. Vector f that minimized graph partitioning minimum cut objective function is given by 2nd smallest eigenvalue  $\lambda_2$  and corresponding eigenvector of the graph Laplacian L. Bi-partition (C1, C2) is expressed as a vector by Eq. 8

$$f_i = \begin{cases} 1, & \text{if } x_i \in C1 \\ -1, & \text{if } x_i \in C2 \end{cases} \tag{8}$$

---

**Algorithm 2: Proposed Clustering Based Movie Recommendation System**


---

**Input :**

Preprocessed User-Item Rating Matrix R.  
 algonames -Names of clustering algorithm in experiment(Kmean, SC-NN, SC-BI-MD, SC-RBF-MD, SC-NN-MD).  
 nc - [2,k] number of clusters.

**Output :**

$C = \{C_1, C_2, \dots, C_k\}$  a set of clusters.  
 $r'_{ij}$  - predicted rating for user i for item j.  
 RMSE, MAE , Precision, Recall, Accuracy and F1 Score for particular number of cluster.

**Begin:**

```

for algoname in algonames do:
  for clustnum in nc do:
    CD ← ClusterData(R, clustnum, algoname)
    for activeUID and activeMID in testset do:
      ClusterID ← findClusterID(CD,activeUID)
      Simusers ← findPeerUser(ClusterID)
      Ratings ← findRatingofActiveM(Simusers,activeMID)
      predRating ← Avg(Ratings)
    RMSE, MAE,Precision, Recall, Accuracy, F1 Score ← MetricCal(trueRating, predRating)

```

**End.**

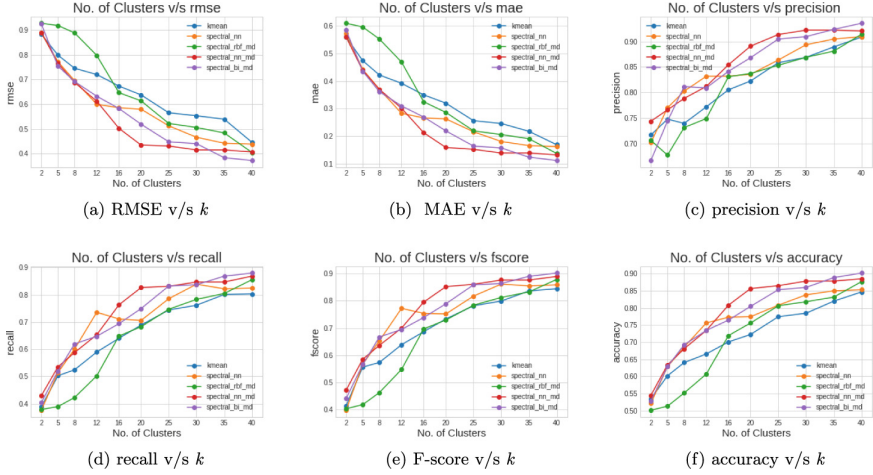

---

## 5 Experimental Results Discussion

We examine SC-BI-MD's performance in the experiment in contrast to Kmean, SC-NN, SC-RBF-MD and SC-NN-MD by using different measures such as RMSE, MAE, precision, recall, accuracy and F-score. Rating predictions were executed on the benchmark MovieLense dataset (ML-Latest-Small) and results were compared. This dataset contains 610 users, 9742 movies, 3683 tag applications and 100836 ratings in a 5-star rating scale. In addition, we have performed a performance study based on the number of clusters versus the RMSE, MAE, precision, recall, accuracy and F-score. All these approaches discussed above are executed on Python Jupiter Notebook.

### 5.1 Results and Discussion

This section presents the discussion of the effectiveness of SC-BI-MD in comparison with K-means, SC-NN, SC-RBF-MD and SC-NN-MD evaluated on the movieLens dataset as per the various estimates, such as RMSE, MAE, precision, recall, accuracy and F-score respectively. The major objective of this section is to show the improvements caused by SC-BI-MD over other traditional methods. The existing approach of K-means clustering Movie Recommendation System [2, 4, 11] and Spectral clustering movie Recommendation System results are represented. We have compared our work with these approaches to show the improvements. In addition, we have presented the results of different clustering algorithms for the Movie Recommendation System to show the improvement caused by our proposed algorithm. The parameter k is used to specify the particular cluster number. In our experiment, the value of k varies from 2 to 40. We



**Fig. 2.** Performance of different clustering algorithms with different number of clusters( $k$ ) on ML-Latest-Small dataset. a RMSE v/s  $k$ . b MAE v/s  $k$ . c precision v/s  $k$ . d recall v/s  $k$ . e F-score v/s  $k$ . f accuracy v/s  $k$ .

took the different values of  $k$  as 2, 5, 8, 12, 16, 20, 25, 30, 35, and 40. The effect of variation in the value  $k$  on different metrics is measured for all six clustering algorithms and presented in Figure.

The RMSE of different algorithms versus the number of clusters is shown in Fig. 2a. Y-axis in Fig. 2a represents RMSE and variation in  $k$  values is presented on X-axis. RMSE denotes the error in prediction. Lower RMSE denotes better prediction. It is analyzed from the figure that the proposed approach SC-BI-MD RMSE is lower than SC-NN, SC-RB-MD and SC-NN-MD. RMSE of K-means is the highest among all six clustering algorithms. It is also found that the RMSE values of all five clustering algorithms are decreasing with an increase in the value of  $k$ . The lowest value of RMSE is 0.3726 for 40 clusters is achieved by SC-BI-MD.

Figure 2b shows the effect of the increase in the value of  $k$  on the MAE of different clustering algorithms. For all algorithms as the number of clusters increases the MAE value decreases. MAE for K-means is the highest for all algorithms and MAE for SC-BI-MD is the lowest. The lowest value of MAE is 0.112, for 40 clusters is achieved by SC-BI-MD.

Figure 2c shows the precision of different algorithms w.r.t. number of clusters. The precision of the proposed algorithm SC-BI-MD is the highest among all algorithms. Precision for SC-BI-MD gradually increases with the number of clusters increases. The highest value of precision is 0.9357, for 40 clusters is achieved by SC-BI-MD. Precision of SC-NN-MD is also comparable with SC-BI-MD.

Figure 2d shows the effect of the increase in the value of  $k$  on the recall of different clustering algorithms. Figure 2d also shows that the recall of SC-BI-MD algorithm is highest. The second highest recall is for algorithm SC-NN-MD. For

16 and 20 clusters, SC-NN-MD recall is more elevated than SC-BI-MD algorithm. The highest value of recall is 0.8796, for 40 clusters, is achieved by SC-BI-MD. On the other hand, the lowest value of recall, 0.803, is achieved by kmeans.

The clustering quality is shown by the F-score. Consequently, we achieve superior clustering here as well. Because of this, we can see that in Fig. 2e, the value of Fscore for SC-BI-MD is the highest for all algorithms when comparing values of F score for different cluster sizes. SC-NN-MD F-score is less than SC-BI-MD but higher than K-means, SC-NN and SC-RBF-MD. The highest value of the F-score is 0.9023, for 40 clusters is achieved by SC-BI-MD. F-score of SC-NN-MD is also comparable with SC-BI-MD.

As Fig. 2f shows the accuracy of SC-BI-MD is comparable with other algorithms. The highest value of accuracy is 0.9013, for 40 clusters is achieved by SC-BI-MD. The accuracy of K-means, SC-NN is less than the accuracy of SC-BI-MD. It is found from Fig. 2f that the accuracy of all algorithms increases in the increase in cluster size.

## 6 Conclusion and Future Work

The Movie Recommender System in the suggested system is constructed utilizing a variety of clustering algorithms. The performance of several clustering methods has been evaluated using the well-known ML-100K data set from MovieLens. We vary the size of clusters from 2 to 40. In this paper, we compare the performance of several clustering methods like Kmeans, Spectral clustering with nearest neighbors and radial Basis Function affinity, Spectral clustering with nearest neighbors and radial Basis Function affinity with Mahalabonis distance, Spectral Bi-clustering with Mahalabonis distance to generate a movie recommendation system. It has been observed that errors like RMSE and MAE are decrease with an increase in cluster size. For the given setup, we find out that our proposed method SC-BI-MD (Spectral Bi-clustering with Mahalabonis distance) achieved a minimum value for RMSE and MAE for cluster size equal to 40. Performance indicators like precision, recall, F1 score and accuracy are increased with an increase in cluster size. For cluster size 40, the highest value of precision, recall, F1 score and accuracy are achieved by SC-BI-MD. The drawback of our strategy is that, while it excels with small datasets, it may have run-time and memory implications with larger datasets. Other nature-inspired algorithms like ANN, PSO, and ACO with Collaborative Filtering can be used to enhance the performance of the suggested system.

## References








1. Agarwal, N., Haque, E., Liu, H., Parsons, L.: Research paper recommender systems: a subspace clustering approach. In: Fan, W., Wu, Z., Yang, J. (eds.) WAIM 2005. LNCS, vol. 3739, pp. 475–491. Springer, Heidelberg (2005). [https://doi.org/10.1007/11563952\\_42](https://doi.org/10.1007/11563952_42)

2. Ahuja, R., Solanki, A., Nayyar, A.: Movie recommender system using k-means clustering and k-nearest neighbor. In: 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 263–268. IEEE (2019)
3. Airen, S., Agrawal, J.: Movie recommender system using k-nearest neighbors variants. *Natl. Acad. Sci. Lett.* 1–8 (2021). <https://doi.org/10.1007/s40009-021-01051-0>
4. Bridge, D., Kelleher, J.: Experiments in sparsity reduction: using clustering in collaborative recommenders. In: O’Neill, M., Sutcliffe, R.F.E., Ryan, C., Eaton, M., Griffith, N.J.L. (eds.) *AICS 2002*. LNCS, vol. 2464, pp. 144–149. Springer, Heidelberg (2002). [https://doi.org/10.1007/3-540-45750-X\\_18](https://doi.org/10.1007/3-540-45750-X_18)
5. De Maesschalck, R., Jouan-Rimbaud, D., Massart, D.L.: The mahalanobis distance. *Chemom. Intell. Lab. Syst.* **50**(1), 1–18 (2000)
6. Dhillon, I.S., Guan, Y., Kulis, B.: A unified view of kernel k-means, spectral clustering and graph cuts. *Citeseer* (2004)
7. Hoseini, E., Hashemi, S., Hamzeh, A.: A levelwise spectral co-clustering algorithm for collaborative filtering. In: *Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication*, pp. 1–6 (2012)
8. Kluger, Y., Basri, R., Chang, J.T., Gerstein, M.: Spectral biclustering of microarray data: coclustering genes and conditions. *Genome Res.* **13**(4), 703–716 (2003)
9. Madeira, S.C., Oliveira, A.L.: Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **1**(1), 24–45 (2004)
10. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: analysis and an algorithm. In: *Advances in Neural Information Processing Systems*, pp. 849–856 (2002)
11. Sarwar, B.M., Karypis, G., Konstan, J., Riedl, J.: Recommender systems for large-scale e-commerce: scalable neighborhood formation using clustering. In: *Proceedings of the Fifth International Conference on Computer and Information Technology*, vol. 1, pp. 291–324 (2002)
12. Von Luxburg, U.: A tutorial on spectral clustering. *Stat. Comput.* **17**(4), 395–416 (2007)
13. Wang, J., Song, H., Zhou, X.: A collaborative filtering recommendation algorithm based on biclustering. In: *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, pp. 803–807. IEEE (2015)
14. Xu, B., Bu, J., Chen, C., Cai, D.: An exploration of improving collaborative recommender systems via user-item subgroups. In: *Proceedings of the 21st International Conference on World Wide Web*, pp. 21–30 (2012)





# Exploring the Use of Bluetooth Low Energy (BLE) Beacons for Enhancing Ecotrails in the Amazon Jungle of Peru

J. Baldeón<sup>1</sup> , D. Auccapuri<sup>1</sup> , A. Masuda<sup>1</sup> , R. Gálvez<sup>1</sup> , E. Díaz<sup>1</sup> ,  
A. Arana<sup>2</sup> , P. Chávez<sup>2</sup>, V. Hernández<sup>3</sup> , and M. Lau<sup>3</sup>

<sup>1</sup> Pontificia Universidad Católica del Perú, Lima, Peru  
{johan.baldeon, dauccapuri, masuda.a, rjgalvezm, ediazm}@pucp.edu.pe

<sup>2</sup> Blueline Advanced Services - Perú, Lima, Peru  
{aarana, info}@bluelineperu.com

<sup>3</sup> Pacaya Samiria Amazon Lodge, Iquitos, Peru  
{victor, maelena}@pacayasamiria.com.pe  
<https://www.pucp.edu.pe/>, <https://www.pacayasamiria.com.pe/>

**Abstract.** The study aimed to design and implement a system that provides visitors of Pacaya Samiria Amazon Lodge Private Reserve in the Amazon Jungle of Peru with real-time location-based information and interactive experiences using their smartphones. The proposed system consists of Bluetooth Low Energy (BLE) beacons placed along the Ecotrails that communicate with visitors' smartphones, providing them with information through AR about the flora, fauna, and cultural heritage. The study utilized a qualitative method approach that involved a focus group. The results suggest that using BLE beacons can significantly enhance visitors' experiences in ecotourism by providing contextual and immersive information and creating a sense of engagement with the natural and cultural surroundings. The study also highlights the importance of designing ecologically sustainable technology solutions, supporting the long-term preservation of the fragile ecosystem of the Amazon Jungle.

**Keywords:** Bluetooth Low Energy · BLE beacons · Augmented Reality · Amazon Jungle · Pacaya Samiria · Ecotourism

## 1 Introduction

Ecotourism plays a crucial role in promoting environmental conservation and supporting local communities. It offers an opportunity for individuals to experience and appreciate the beauty and significance of natural ecosystems. However, traditional methods of guiding visitors, such as signage and physical guides, often lack interactivity and fail to provide contextual information in real-time [5, 7, 14]. In this context, the integration of emerging technologies becomes essential in creating engaging and immersive experiences for ecotourists.

Supported by ProInnovate Perú. PIEC2-7-P-191-22.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024  
F. P. García Márquez et al. (Eds.): ICCIDA 2023, SCI 1145, pp. 334–343, 2024.  
[https://doi.org/10.1007/978-3-031-53717-2\\_32](https://doi.org/10.1007/978-3-031-53717-2_32)

In recent years, there has been a growing interest in integrating innovative technologies to enhance the tourist experience while promoting sustainability in ecotourism destinations [11, 13]. One such technology that shows great potential is Bluetooth Low Energy (BLE) beacon technology [3, 8–10, 15].

This paper aims to explore the use of BLE beacons for enhancing Ecotrails in the Amazon Jungle of Peru, specifically in the Pacaya Samiria Amazon Lodge Private Reserve [1]. These beacons act as proximity-based triggers that communicate with visitors' smartphones. By leveraging augmented reality (AR) technology, visitors receive real-time information about the local flora, fauna, and cultural heritage, enriching their understanding and appreciation of the ecosystem. The primary contribution of this work lies in the design and implementation of a system that enhances visitors' experiences in ecotourism by utilizing BLE beacons and smartphones. The proposed system provides real-time location-based information and interactive experiences through AR, fostering a deeper engagement with the natural and cultural surroundings.

To achieve the objectives of the study, a qualitative research approach was adopted, involving a focus group of participants who experienced the BLE beacon-based system during their visit to the Pacaya Samiria Amazon Lodge Private Reserve. The focus group discussions aimed to gather insights and feedback on the effectiveness and impact of the system on the visitors' experiences.

The study reveals that the use of BLE beacons significantly enhances visitors' experiences by providing contextual and immersive information. Visitors reported a heightened sense of connection and appreciation for the ecological and cultural aspects of the Amazon Jungle. By providing real-time, context-specific information through AR, the system creates a more immersive and engaging experience, enabling visitors to develop a deeper connection with the natural and cultural surroundings. The interactive nature of the system fosters a sense of curiosity and exploration, making the Ecotrails more informative and enjoyable for the visitors.

Furthermore, the study highlights the importance of designing ecologically sustainable technology solutions. By opting for BLE beacons instead of conventional signage or physical guides, we aim to minimize the impact on the fragile ecosystem of the Amazon Jungle. This consideration aligns with the long-term preservation and sustainability goals of ecotourism practices.

## 2 Literature Review

Ecotourism, sustainability, and the integration of technology in tourism have been topics of increasing interest in recent years [6]. This section reviews the relevant literature on these areas, explores previous studies that have utilized Bluetooth Low Energy (BLE) beacons or similar technologies to enhance visitor experiences in ecotourism or cultural heritage sites and identifies the gaps and limitations in the existing literature that this study aims to address.

**Ecotourism and Sustainability.** Ecotourism, defined as responsible travel to natural areas that conserves the environment and improves the well-being of local communities, has gained prominence as a sustainable approach to tourism [5]. Many studies have highlighted the benefits of ecotourism in terms of promoting environmental conservation, supporting local communities, and providing immersive and educational experiences for visitors [7,14]. However, the effective implementation of ecotourism practices requires innovative approaches that balance visitor experiences with the preservation of natural environments [2].

**Technology in Tourism.** Technology has increasingly been integrated into the tourism industry to enhance visitor experiences and support sustainable practices [11,13]. Various technologies, such as augmented reality (AR), mobile applications, and location-based services, have been explored in the context of tourism [12]. These technologies offer opportunities to provide real-time information, interactive experiences, and personalized recommendations to visitors, ultimately enhancing their engagement and satisfaction.

**BLE Beacons in Ecotourism and Cultural Heritage Sites.** In recent years, BLE beacons have emerged as a promising technology for enhancing visitor experiences in ecotourism and cultural heritage sites [3,4,8–10,15]. BLE beacons are small, low-power devices that can transmit signals to nearby smartphones, enabling location-based services and interactions. Previous studies have demonstrated the potential of BLE beacons in providing contextual and immersive information to visitors, improving navigation, and facilitating engagement with the surroundings.

**Gaps and Limitations in the Existing Literature.** While previous studies have highlighted the potential of BLE beacons and similar technologies in enhancing visitor experiences in ecotourism and cultural heritage sites, there are several gaps and limitations in the existing literature that this study aims to address. Firstly, there is a lack of research specifically focusing on the use of BLE beacons in the context of ecotrails in the Amazon Jungle of Peru. This study seeks to fill this gap by exploring the application of BLE beacons in this unique and biodiverse setting. Secondly, while previous studies have demonstrated the benefits of BLE beacons, there is limited research on the potential challenges and implications associated with their implementation. This study aims to shed light on practical considerations, such as beacon placement, communication protocols, and user interface design, to provide insights for future implementations in ecotourism contexts.

Lastly, there is a need to emphasize the importance of designing ecologically sustainable technology solutions in the context of ecotourism. While technology can enhance visitor experiences, it should also support the long-term preservation of the fragile ecosystem of the Amazon Jungle. This study aims to address this aspect by considering sustainability as a crucial factor in the design and implementation of the BLE beacon system.

Table 1 provides an overview of the key features and differences between the proposed platform that utilizes BLE beacons and state-of-the-art approaches such as QR codes and GPS-based systems. It highlights the advantages of the proposed platform in terms of technology used, system architecture, interaction method, real-time location-based information, information delivery through AR, interactivity, availability of flora and fauna information, cultural heritage information, sense of engagement with surroundings, and ecological sustainability consideration.

**Table 1.** Comparative table highlighting the differences between the proposed platform that uses BLE beacons and the state-of-the-art approaches

Features	Proposed Platform	State-of-the-Art Approaches
Technology Used	BLE Beacons	QR Codes, GPS
System Architecture	Beacon-Smartphone	QR Code-Scanner, GPS
Interaction Method	Proximity-based	Scan QR Codes, Location-based
Real-time Location-based Inform	Yes	Limited or No
Information Delivery	Augmented Reality	Text, Images, Audio
Interactivity	Interactive AR	Static Information
Flora and Fauna Information	Yes	Limited or No
Cultural Heritage Information	Yes	Limited or No
Sense of Engagement with Surroundings	High	Moderate or Low
Eco Sustainability Consideration	Yes	Varies

### 3 Methodology

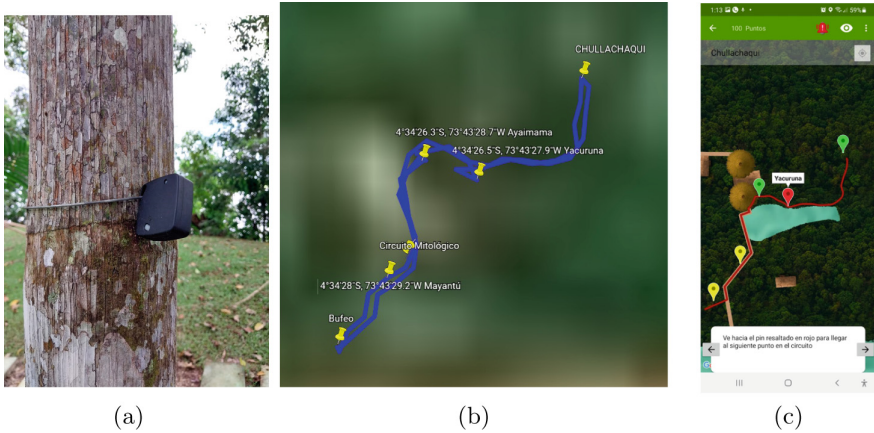
#### 3.1 Overview of the Method and Proposed Approach

This study employs a qualitative research approach to explore the use of Bluetooth Low Energy (BLE) beacons for enhancing Ecotrails in the Amazon Jungle of Peru. Qualitative research allows for an in-depth understanding of participants' experiences, perceptions, and interactions with the BLE beacon system. A focus group method is utilized as it facilitates group discussions and enables the exploration of multiple perspectives on the topic.

#### 3.2 Designing and Implementing the BLE Beacon System

The design and implementation of the BLE beacon system involved several steps. First, we conducted a thorough analysis of the Ecotrails in the Amazon Jungle

to identify suitable locations for placing the BLE beacons. Factors such as trail accessibility, biodiversity hotspots, and significant cultural heritage sites were considered. Figure 1a shows a beacon installed on an Ecotrail. Figure 1b shows the map of the Mythological Ecotrail and the position of each of the BLE beacons and Fig. 1c shows the Mythological Ecotrail map in the “Pacaya Samiria” AR App.



**Fig. 1.** (a) Beacon installed on an Ecotrail. (b) Definition of the Mythological Ecotrail map using a GPS tracker. (c) Mythological Ecotrail Map in the “Pacaya Samiria” AR App.

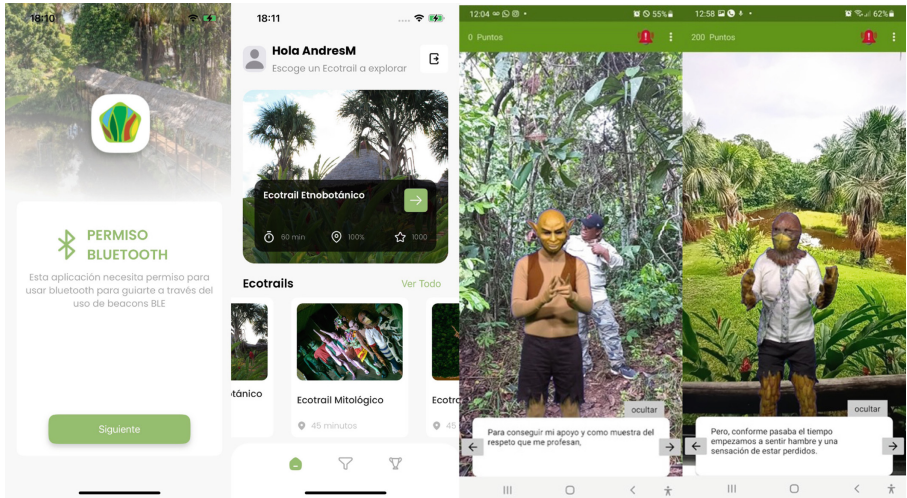
Next, a communication protocol was developed to ensure seamless interaction between the BLE beacons and visitors’ smartphones. This protocol enabled transmitting location-based information and interactive experiences through augmented reality (AR) technology. The user interface on smartphones was designed to provide a user-friendly and intuitive experience for visitors. It included features such as real-time trail navigation, information about flora and fauna, cultural heritage insights, and interactive AR elements (see Fig. 2).

**3.3 Participant Recruitment and Selection**

Participants for the study were recruited from visitors to the Amazon Jungle Ecotrails who had smartphones capable of interacting with BLE technology. A purposive sampling technique was employed to select participants with prior ecotourism experience and who were willing to engage in focus group discussions.

**3.4 Data Collection Process**

Data collection involved conducting focus group sessions with the selected participants. The sessions were held in a suitable location within the Pacaya Samiria



**Fig. 2.** GUI of “Pacaya Samiria” AR Application.

Amazon Lodge Private Reserve in the Amazon Jungle, ensuring a natural and comfortable environment. We facilitated group discussions by presenting scenarios and prompts related to visitors’ experiences with the BLE beacon system. The participants were encouraged to share their opinions, suggestions, and any challenges they encountered during their interaction with the system. Audio and video recordings were made to capture the discussions, ensuring accurate documentation of participants’ responses and non-verbal cues. Field notes were also taken to record observations and contextual details.

### 3.5 Data Analysis

The data obtained from the focus group discussions were transcribed and organized for analysis. A thematic analysis approach was employed to identify recurring themes, patterns, and insights related to visitors’ experiences with the BLE beacon system. We immersed ourselves in the data, identifying codes, categories, and sub-themes from the discussions. The analysis involved a systematic process of coding, categorizing, and interpreting the data to derive meaningful findings. Through this methodology, the study aims to comprehensively understand the visitors’ perspectives and experiences with the BLE beacon system in enhancing Ecotrails in the Pacaya Samiria Amazon Lodge Private Reserve in the Amazon Jungle of Peru. The qualitative approach and focus group method provided us with valuable insights into the usability, effectiveness, and potential improvements of the technology in ecotourism settings.

## 4 Results

The findings of this study provide valuable insights into the use of BLE beacons for enhancing Ecotrails. The results are based on the analysis of focus group discussions with participants who interacted with the BLE beacon system during their Ecotrail experiences.

Several key themes and patterns emerged from the data, shedding light on the participants' feedback and reactions to the BLE beacon system.

**Benefits of Using BLE Beacons for Ecotrails.** Participants highlighted various benefits associated with the use of BLE beacons in ecotourism settings. One significant advantage was the provision of contextual and immersive information about the flora, fauna, and cultural heritage of the Amazon Jungle. The interactive experiences facilitated by augmented reality (AR) technology were particularly appreciated, as they allowed visitors to engage with their surroundings in a more meaningful and educational way. Participants expressed a sense of wonder and fascination as they discovered hidden details and gained a deeper understanding of the natural and cultural significance of the Ecotrails. Additionally, the real-time location-based information provided by the BLE beacon system proved beneficial for trail navigation and safety. Visitors reported feeling more confident and secure in their exploration of the Ecotrails, knowing that they had access to accurate and up-to-date guidance throughout their journey.

**Challenges and Considerations.** Despite the overall positive feedback, some challenges and considerations were identified by the participants. Connectivity issues were reported in certain areas of the Ecotrails, affecting the reliability of the BLE beacon system. This highlighted the need for a robust and stable network infrastructure to ensure uninterrupted communication between the beacons and visitors' smartphones.

**Impact on Visitors' Experiences and Engagement.** The findings revealed a significant impact of the BLE beacon system on visitors' experiences and engagement with the natural and cultural surroundings of the Amazon Jungle. Participants expressed a heightened sense of connection and appreciation for the environment as they interacted with the system. The immersive and interactive features provided by the BLE beacons fostered a deeper understanding of the ecosystem and cultural heritage, enhancing the overall ecotourism experience. The use of augmented reality (AR) technology played a vital role in capturing visitors' attention and creating memorable moments. It sparked curiosity, encouraged exploration, and generated a sense of awe and excitement among participants.

**Implications for Sustainability and Conservation.** The study findings have significant implications for sustainability and conservation efforts in ecotourism. Participants recognized the potential of the BLE beacon system to raise

awareness and educate visitors about the importance of preserving the fragile ecosystem of the Pacaya Samiria in the Amazon Jungle. By fostering a deeper connection with nature and cultural heritage, the system encouraged visitors to become more conscious of their impact and actively engage in sustainable practices. The study also highlighted the importance of considering the ecological footprint of technology solutions in ecotourism. Participants stressed the need for responsible and environmentally friendly design principles to ensure the long-term preservation of the Amazon Jungle's biodiversity and natural resources.

## 5 Discussion

The study's findings reveal that the use of BLE beacons significantly enhances visitors' experiences in ecotourism by providing contextual and immersive information about the flora, fauna, and cultural heritage of the Amazon Jungle. The interactive experiences facilitated by AR technology deepen understanding and appreciation of the natural and cultural surroundings, resulting in a more meaningful and educational ecotourism experience.

Theoretical implications of this work advance our understanding of how technology can enhance visitor experiences and engagement in ecotourism settings. BLE beacons and AR technology introduce a new dimension to ecotourism, offering interactive and immersive experiences beyond traditional information dissemination.

Practically, this study highlights the potential of BLE beacons as valuable tools for ecotourism destinations like the Amazon Jungle of Peru. The system provides real-time location-based information, enhances trail navigation, and offers engaging and educational content about the area's natural and cultural heritage. By incorporating technology sustainably and responsibly, ecotourism destinations can create unique and memorable experiences while fostering environmental stewardship.

A strength of this study is its focus on the Pacaya Samiria Amazon Lodge Private Reserve, an ecologically significant location. Conducting the research there provide valuable insights into the application of BLE beacons for enhancing Ecotrails in a biodiverse and culturally rich environment. The qualitative approach and focus groups offer an in-depth exploration of participants' experiences, yielding rich and nuanced data.

Limitations to consider include the small sample size of participants from a specific location, limiting generalizability. Future research could include a larger and more diverse sample for broader representation. Additionally, incorporating perspectives from local communities, ecotourism operators, and environmental stakeholders would provide a comprehensive understanding of the potential impacts of the BLE beacon system in the Amazon Jungle.

## 6 Conclusion

In this study, the use of BLE beacons for enhancing Ecotrails in the Pacaya Samiria Amazon Lodge Private Reserve in the Amazon Jungle of Peru was



explored. The findings emphasize the significant potential of this technology in enhancing visitors' experiences, fostering engagement with the natural and cultural surroundings, and promoting sustainability and conservation in ecotourism settings. Through qualitative research and focus group discussions, it was found that the implementation of BLE beacons along the Ecotrails provided visitors with contextual and immersive information about the flora, fauna, and cultural heritage of the ecolodge reserve in the Amazon Jungle. The interactive experiences facilitated by augmented reality (AR) technology fostered a deeper understanding and connection with the environment, resulting in a more meaningful and educational ecotourism experience. Participants reported increased engagement and appreciation of the natural and cultural surroundings, indicating the positive impact of BLE beacons on visitors' experiences.

The main contribution of this work is the design and implementation of a system that enhances visitors' experiences in ecotourism by providing real-time location-based information and interactive experiences through BLE beacons and smartphones. Furthermore, this study contributes to the existing body of knowledge by exploring the application of BLE beacons in ecotrails in the Amazon Jungle of Peru, emphasizing the importance of designing ecologically sustainable technology solutions and bridging the gap between sustainability, tourism, and technology. The practical implications of the study's findings extend beyond the Amazon Jungle of Peru. Other ecotourism destinations with fragile ecosystems and rich cultural heritage can also benefit from integrating BLE beacons and AR technology into their trail systems. These technologies have the potential to enhance visitor experiences, increase awareness and understanding of the natural and cultural surroundings, and promote sustainable practices.

Based on the findings, several avenues for future research are recommended. The long-term effects of using BLE beacons on visitors' behaviours and attitudes towards sustainability and conservation should be investigated. Engaging local communities and indigenous knowledge in the design and implementation of technology solutions would contribute to more culturally sensitive and inclusive approaches. Additionally, addressing technical challenges associated with connectivity in remote and ecologically sensitive areas is crucial.

**Acknowledgements.** We thank project PIEC2-7-P-191-22 (ProInnovate Perú).

## References

1. Pacaya Samiria Amazon Lodge. <https://www.pacayasamiria.com.pe/>. Accessed 02 May 2023
2. Abdelli, M.E.A., Mansour, N., Akbaba, A., Serradell-Lopez, E.: Sustainability, Big Data, and Corporate Social Responsibility. CRC Press, Boca Raton (2022). <https://doi.org/10.1201/9781003138051>
3. Belka, R., Deniziak, R.S., Łukawski, G., Pięta, P.: BLE-based indoor tracking system with overlapping-resistant IoT solution for tourism applications. *Sensors* **21**(2), 329 (2021). <https://doi.org/10.3390/s21020329>

4. Cardoso, A.F.S., Sousa, B.B., da Cunha, A.C.G.: Mobile applications in urban ecotourism: promoting digitization and competitive differentiation, pp. 349–369. Springer, Cham (2022). [https://doi.org/10.1007/978-3-030-97877-8\\_10](https://doi.org/10.1007/978-3-030-97877-8_10)
5. Conway, T., Cawley, M.: Defining ecotourism: evidence of provider perspectives from an emerging area. *J. Ecotour.* **15**(2), 122–138 (2016). <https://doi.org/10.1080/14724049.2016.1153105>
6. Cosio, L.D., Buruk, O.O., Fernández Galeote, D., Bosman, I.D.V., Hamari, J.: Virtual and augmented reality for environmental sustainability: a systematic review. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. CHI 2023, Association for Computing Machinery, New York, NY, USA (2023). <https://doi.org/10.1145/3544548.3581147>
7. Font, X., Sanabria, R., Skinner, E.: Sustainable tourism and ecotourism certification: raising standards and benefits. *J. Ecotour.* **2**(3), 213–218 (2003). <https://doi.org/10.1080/14724040308668145>
8. Gala, A., Borgaonkar, C., Kulkarni, V., Wakode, M., Kale, G.: Contextual flow of information in tourism using BLE proximity detection to enhance the tourism experience. In: 2023 International Conference on Emerging Smart Computing and Informatics (ESCI), pp. 1–6 (2023). <https://doi.org/10.1109/ESCI56872.2023.10100063>
9. Gonçalves, F., Martins, A.L., Ferreira, J.C., Marques, E., Andrade, M., Mota, L.: Tourism guidance tracking and safety platform. In: Martins, A.L., Ferreira, J.C., Kocian, A. (eds.) INTSYS 2019. LNICST, vol. 310, pp. 162–171. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-38822-5\\_11](https://doi.org/10.1007/978-3-030-38822-5_11)
10. Hiramatsu, Y., Ito, A., Sasaki, A.: Developing an application in the forest for new tourism post COVID-19 -experiments in Oku-Nikko national park. In: Moreno-Díaz, R., Pichler, F., Quesada-Arencibia, A. (eds.) EUROCAST 2022. LNCS, vol. 13789, pp. 250–257. Springer, Cham (2022). [https://doi.org/10.1007/978-3-031-25312-6\\_29](https://doi.org/10.1007/978-3-031-25312-6_29)
11. Kurnaz, H.A., Ön, F., Yüksel, F.: Sustainability, Big Data, and Corporate Social Responsibility, Chapter A New Age in Tourist Guiding: Digital Tourism and Sustainability. CRC Press, Boca Raton (2022). <https://doi.org/10.1201/9781003138051>
12. Liang, L.J., Elliot, S.: A systematic review of augmented reality tourism research: what is now and what is next? *Tourism Hospitality Res.* **21**, 15–30 (2021). <https://doi.org/10.1177/1467358420941913>
13. Pavlidis, G., et al.: Sustainable ecotourism through cutting-edge technologies. *Sustainability* **14**(2), 800 (2022). <https://doi.org/10.3390/su14020800>
14. Stem, C.J., Lassoie, J.P., Lee, D.R., Deshler, D.D., Schelhas, J.W.: Community participation in ecotourism benefits: the link to conservation practices and perspectives. *Soc. Nat. Res.* **16**(5), 387–413 (2003). <https://doi.org/10.1080/08941920309177>
15. Vo, V.A., Nguyen, D.V., Tran, T.T., Pham, M.N.N., Le, T.D., Vo, P.L.: A tourism support framework using beacons technology. In: 2021 8th NAFOSTED Conference on Information and Computer Science (NICS), pp. 342–347 (2021). <https://doi.org/10.1109/NICS54270.2021.9701537>



# Identifying Success Factors in Business Intelligence Using Resource-Based Approach: A Literature Review

Ruksana Banu<sup>(✉)</sup> 

Muscat College, P.C 112, P.O. Box 2910, Muscat, Sultanate of Oman  
ruksana@muscatcollege.edu.om

**Abstract.** In recent years, business intelligence has been a mechanism that provides important decision-making information to ensure firms' sustainability and add value to stakeholders. This study provides a review of literature that investigates how and why firms can use resource-based approach to identify factors for business intelligence success. A qualitative approach was adopted to carry out a content analysis of existing literature. The findings indicated that factors like firm's tangible and intangible resources are associated with information technology and organizations' system, which supports enhancing firm's business intelligence capabilities and performance. Based on the literature review findings, a conceptual model has been developed to be empirically tested in future studies. The consolidated findings may contribute to existing literature on business intelligence and resource-based approach may be used in identifying a firm's success factors in the context of business intelligence.

**Keywords:** business intelligence · resource-based view · tangible resources · intangible resource · information technology · capabilities

## 1 Introduction

Many businesses use Business Intelligence (BI) technologies to obtain a competitive advantage. Through systematic data capture, collection, analysis, interpretation, and utilization, BI elements assist firms in understanding both internal and external business operations in a better manner. It is claimed that BI enables firms to discover opportunities and risks that may arise in the market and understand the changing behavioural pattern of consumers, suppliers, and competitors. According to the Mordor intelligence market overview, the BI market was valued at around 20.5 billion USD in 2020. By 2026, it is expected to increase to 40.5 billion USD with a CAGR growth rate of 12% between 2020–26 [1]. The utilization of BI and analytic tools has grown steadily because contemporary firms use BI dashboards and performance scorecards to examine their business metrics and to understand critical business performance indicators. The usage of BI has emerged due to advancements in analytic tools and technology, as it enhances firms' capability to conclude valuable insights on main business drivers. BI potential and benefits to firms range from optimizing the internal business process to obtaining a competitive advantage

in the market. Firms use BI to accelerate and improve the decision-making process and identify main business problems and market trends that need to be focused on and addressed [2]. During the pandemic, BI usage increased since it enabled firms to predict demand and identify supply-chain disruptions. This supports firms in determining the crisis intervention strategies and resource management factors. Even though business intelligence is one of the widely purchased technologies, the applications of BI may fail, or firms need to gain more knowledge to make appropriate use of the BI [3, 4]. Moreover, the reasons for failure or gaining more knowledge are unclear, and the success factors of BI are yet to be thoroughly examined. Thus, there is a need to examine if a resource-based approach and theory can bridge this gap and provide a theoretical foundation for BI fields of study.

BI capabilities are vital features that can assist firms in improving their performance, adapting to changing business environments, and improving organizational functionalities. Using a dynamic capabilities viewpoint, this study aims to shed further light on the concept of BI via a resource-based view (RBV). As a result, the key objectives of this study are 1) to examine the concepts and success factors concerned with BI and 2) to develop a conceptual model of BI based on a resource-based approach. This article is presented in three sections to address the objectives of this study. The first section provides concepts and theoretical approaches to BI and resource-based view. The second section offers critical success factors associated with BI through a resource-based approach, and the third section provides a conceptual model of BI from RBV.

## 2 Research Method

This study aimed to investigate the essential success elements linked with business intelligence from a resource-based approach. The study examined whether existing success factors connected with the development of each of these systems might be used in implementing BI systems as an extension of a resource and capabilities. The goal was to see if the resource-based view was relevant and if any BI success factors still needed to be recorded in the academic literature. The research question “What are the success factors of BI from a resource-based approach?” was utilized to lead the research to attain the research goal. The study used qualitative approaches to investigate the essential success variables linked with BI adoption. Qualitative research methodologies aim to “provide knowledge of the heterogeneous evidence and information system”. This study assumed that resource-based views and business intelligence success factors are integrated based on simplistic assumptions”. In the conceptual review of the literature, a qualitative research methodology has been proven to be an effective method of investigation. A content analysis strategy was proposed as part of the methodology to confirm and find new success criteria relevant to BI adoption from a resource-based view. Content analysis can be considered a method for realist review that examines studies to identify themes and contexts of interest. In content analysis, documents collected during research are reviewed and summarized.

### 3 Theoretical Approach

The theoretical approach is essential, as it supports understanding the concepts and helps to examine as well as identify the key variables and aspects associated with business intelligence and resource-based view.

#### 3.1 Resource-Based Approach

In the compensatory business environment, the business intelligence topic has gained much attention as it enables the organization to develop strategies that support explanations and business development. The resource-based approach claims that disparities in business performance can be explained when stretched resources are economically valued (exploiting opportunities and neutralizing threats) and relatively support to retain competitive advantage, as well as facilitating mobilizes of resources within and outside the firms. The resource-based view has been used in various literature to explain business value, including technological advancement. The resource-based theory supports firms in determining the firm-specific sets of resources for competitive and comparative intelligence [5]. As described in resource-based theory, an organization's strategies determine its resources and capabilities, which results in developing its ability to set essential competencies, develop a competitive edge and create value. Moreover, supports acquiring, configuring, re-configuring, and expanding available resources [6]. For instance, to offer a sustained competitive advantage, the firm's resources must be able to meet VRIN criteria, that is Valuable (enables the organization to devise a value-creating strategy), Rare (offer services and products that are short in supply), Inimitable (products/services that competitors cannot easily duplicate) and non-substitutable (a competitor cannot replace the product/services). In a broader sense, resource-based aspects are associated with resources like intangible categories such as organizational, human, and network resources [5, 7]. Accordingly, resource-based knowledge allows firms to acquire, access, and preserve intangible assets, since these resources function as means to combine and transform tangible inputs. Studies have indicated that BI technology, like other ICT, does not meet the VRIN criteria. But it can be coupled synergistically with current organizational resources to create new VRIN resources [3, 6]. Earlier studies have focused on the antecedents or situations under which the value system contributes to business success. Business intelligence is particularly beneficial in utilizing the firm's capabilities and other driving factors are the organizational resources and value chain systems. This study uses the resource-based approach as a frame of reference to determine business intelligence success factors and how business intelligence capabilities contribute to creating and sustaining business value.

#### 3.2 Business Intelligence (BI)

Business intelligence (BI) is a broad concept of tools and techniques for obtaining, storing, cleansing, analysing, and delivering data to organization users to help make improved business choices in less time. According to Stackowiak, et al. [8], business intelligence is the way of acquiring large amounts of information, analysing it, and portraying a high of reports which compress the significance of that data into the framework

of business actions, enabling management to make essential daily operations decisions. Employees, consultants, customers, suppliers, and the public can access BI information and services [5, 6]. Users must rely on accurate information to make critical decisions at the right moment. Business users can extract meaningful information using data analysis, reporting, and query tools. BI has made it simple to manage and acquire a wide range of data due to the necessity of information and proper knowledge of market conditions. Before the information age, accessing any information was extremely difficult due to the absence of computer methods, so most business decisions were based on intuition [5, 9]. Business intelligence is becoming increasingly common in virtually all organizations in today's globalized world. Thanks to technological advancements, it has become easier to gather large amounts of data within a short period, and it has become easier to evaluate the data for long-term strategic decisions due to computing resources. In contrast, BI's history shows that it is a relatively new concept designed to assist business decision-making. Business intelligence enables firms to align their key performance indicators (KPIs) with objectives. Firms develop key performance indicators (KPIs) according to organizational systems and strategies; this helps the firms to set the competitive advantage factors [3, 10, 11]. However, the data should be available in accurate and easily accessible formats so that the setting of KPIs and other strategic decisions can be made at the right time and in the right way. Overall, early commercial programs used their databases to serve their functions. Such databases became "information islands", and more business processes were automated with the advancement of information technology [3, 5]. Various systems, such as sales, accounting, production, and human resources, can provide valuable insights into corporate operations, providing historical, current, and predictive views [9]. Thus, it can be implied that a business intelligence system refers to the process and techniques that enable a firm to simplify and utilize information so that decision-makers can make business decisions more efficiently and efficiently.

#### **4 Resource-Based Approach towards BI Success Factors**

The BI Success Model developed by Wixom and Watson [12] describes the need for more strategic factors that contribute to the success of BI projects. At the same time, a few authors have stressed the importance of organizational alignment between business objectives and firms' approaches for the success of a BI initiative [9]. In the literature study, many of the success factors for BI implementation are not specified in the context of the BI environment. The success characteristics are recognized in terms of data usage for customer relationship management, data storage for knowledge management, use of intangible sources or value chain aspects in supply chain management, and resource capabilities to gain and use geographic information systems. They can be applied to other IS initiatives [5]. Thus, it implies that there is a necessity to integrate data from diverse source systems is, nevertheless, a success component that is uniquely specific to BI. From the content analysis of the literature study on BI and RBV following critical success factors may be helpful for firms to consider gaining maximum benefit from BI adoption.

#### 4.1 Intangible and Tangible Resources

The phrases “resources” and “capabilities” have been used interchangeably. The research-based view (RBV) or approach refers to resources used as inputs into a firm’s production process, such as IT equipment. In contrast, capabilities refer to a firm’s ability to exploit IT equipment (resources) through organizational processes [9]. Few studies discussed that tangible and intangible resources have an impact on corporate capabilities, and business intelligence supports firms to improve those capabilities, further increasing organizational productivity [3, 10]. Few studies have indicated that strategic option like integrating IT with organizations’ systems is vital, as it supports adopting and using BI in the decision-making process and creates an opportunity for developing an innovative mindset among employees [3]. Furthermore, an integrated system within the decision-making process ensures the ethical usage of data and enhances value-driven activities and capabilities. BI initiatives are associated with knowledge intensive. The main advantage is that it facilitates in managing technical and business aspects, and improves enterprising knowledge and skills [13]. However, the organization’s assumptions, values, and behavioural aspects create culture, often referred to as the “personality of the organization”, and these are developed through time [3, 6]. The firm’s strategic goal affects business information and business intelligence usage. For instance, the value system benefits in generating firms’ portfolios and develops access to resources and skills. As a result, adopting business intelligence is perceived to be an effective tool for improving the firm’s strategic goal, performance, and productivity [14]. But a disadvantage may appear if the usage and adoption are not identified or lacks necessary resources and skills. The firm’s capabilities become more difficult for competitors to understand and imitate over a period, particularly when the usage of BI is accelerated [15]. Much research discusses the link between various resources and a firm’s performance, where business intelligence mediates. This is because BI is viewed as IT tangible resources (IT infrastructure) which is a vulnerable source of value and gaining competitive advantage. This is because competitors may find it difficult to imitate intangible resources such as business intelligence [16]. Among BI activities, technology refers to developing and using hardware, software, and data as components [3]. Key success factors are often associated with managing a high-quality integrated data resource and integrating BI systems with ERP systems. Moreover preparing data through reporting systems and usage of advanced statistical analysis tools to discover patterns, predict trends, and optimize business processes are just a few ways that firms can improve business operations [10, 15]. This implies that adopting BI results in a firm’s capabilities and through the effective use of resources they can sustain businesses.

#### 4.2 Enterprise Resource Planning (ERP)

The macro-level aspects support identifying success factors and studies have conferred that any specific industry relies on three to six success or failure factors [17]. For instance, an ERP (Enterprise resource planning) system can be used to integrate heterogeneous data sources for better data quality. BI has been deployed as an extension of firms’ ERP systems [18]. The ERP system delivers business operations and transaction data, whereas BI allows for analyzing this data to improve performance [19]. The essential

elements of BI have been implemented on standalone systems, but this is different in an ERP environment, where BI is closely related to other parts of the business process [20]. Many of the success factors described in the literature are connected to the adoption and success of BI depends up on the implementation of ERP systems. For instance, from a RBV the advantage of ERP is that it enables the organization to develop a creative and innovative culture among the employees, which may provide a source to gain or sustain a competitive advantage in the market. There has been substantial research into discovering essential success elements for ERP systems. But limited studies have discussed BI systems as extensions of ERP systems [17, 18]. The key disadvantage is that the success characteristics of ERP systems may not be automatically applied. Henceforth, there is a need for alternatives like BI systems to explore resources that can enhance firms' capabilities. Some research discussed installing BI systems effectively in ERP environments, particularly in the data storage arena that underpins the BI system's functionality [21]. Based on practitioners' shared experiences and strategies for BI success factors literature, it can be observed that an ERP system is a vital tool in strategic management, and BI factors can be enhanced by creating an effective ERP system.

### 4.3 Data: Process, Storage, and Usages

Business intelligence success is primarily influenced by management support and the proper use of data sources and information [22, 23]. For instance, an effective database storage strategy must be easy to navigate and provide helpful information that supports firms in making and implementing quick decisions. Thus, user satisfaction and end-user skills are essential for the success of data storage. A business-driven approach should be embedded with adequate resources, budgetary support, and skills to use data in decision-making [5, 24]. In a few studies, it has been suggested that data quality, a flexible enterprise model, data governance, automated data extraction methods, and hardware/software systems are some factors that can ensure successful BI implementation [25]. The quantity and kind of resource systems depend on the correct data usage. For instance, data integration is vital in creating firms' BI capacity. It is evident that the diversity and number of source systems directly impact this success factor. Study results have shown that data quality and system quality directly affect the success of data storage, with system quality playing a vital role [22, 26]. BI adoption and success factors are also influenced by the project team's skills, managerial support, adequate resources, and user involvement.

### 4.4 Key Findings

Looking at data taxonomy, the factors like volume, diversity, and velocity of big data are the main features [10, 27]. However, data quality and suitability are also vital for firms to increase their dynamic capabilities. Thus, few studies have discussed that data has become a significant resource within the firm. The McKinsey analysis classified data as crucial as traditional tangible assets like capital, labour, and commercial goods [28]. As a result, data ownership can be regarded as a critical tangible asset along with intangible assets (business infrastructure, IT systems). The firm's intangible assets govern the collection of data and its optimal usage, and this necessitates keeping up with



knowledge and skills, which the firm may only sometimes be able to do [3]. Thus, the role of data analytics is crucial as it may support firms in understanding the significance of effective data management [5]. For example, with the increase in the volume of data in the organization, the firms need to set the framework for data management with detailed structures and activities. This can further increase the dynamic capacities of the firm and comprehend the better use of BI to operate the business activities cost-effectively. Firms should recognize that data-driven culture is critical to the company’s long-term success [25, 29]. The firm’s dynamic capabilities are where the culture of ample data use is embraced by top management down to the individual worker. Moreover, humans’ knowledge, skills, and abilities should be recognized. A firm’s organizational structure and system should include resource managers, programmers, cloud-services managers, and so on to create key success factors for competitive advantage using competencies, human resources, and BI features.

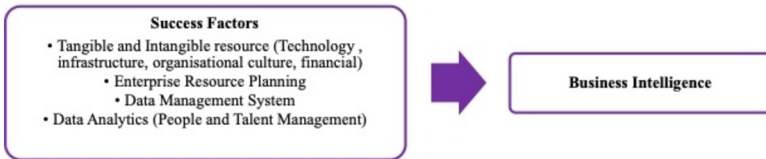
### 5 Summary of Findings and Future Research

From the content analyses of the literature study on BI and its success factors from RBV, Table 1 was developed by the researcher to provide a summary of resource and capabilities (R&C) analysis concerning BI success, and Fig. 1 provides a conceptual model which can be empirically tested via quantitative or qualitative data. An organization’s financial capabilities can be enhanced if it develops a system to control how it uses BI resources and invests in BI initiatives. To respond to changing business environments, BI resources and organizational capacities must be replenished regularly. This will also

**Table 1.** Summary of R&C Analysis Concerning BI Success (Source: Developed by the Researcher)

Factors	Resources		Capabilities	Source
Intangible	Technology	Strategic options associated with technology and integration of IT with organizational system	Data Management System (Storage and Usage)	Bayrak [27]; Mikalef, et al. [10]; Chatterjee, et al. [3]
	Organizational Culture	Creating and developing an innovative mindset and ethical usage of data. Decision-making process	Value-driven, ERP	Appelbaum, et al. [17]; Khan and Quadri [13]
Tangible	Infrastructure	IT hardware and software	Investment, IT Adoption	Işık, et al. [30]; Rao and Provodnikova [5]
	Financial	Allocation of budgets for IT aspects		
Human	People	IT and ICT skills, managerial and leadership skills	Stakeholder Management, Data Process, Data Analytics, ERP	Beck and Wiersema [31]; Teece [9]; Mikalef, et al. [10]; Rao and Provodnikova [5]
	Talent Management	Seamline the use of IT systems and software. Updated knowledge of IT applications and so on		

support mitigating the resistance to change. Gathering, selecting, aggregating, analysing, and disseminating information are all part of the process, and most of these operations are under the purview of the BI aspects. Internal and external processes are the two types of processes, where accounting, finance, manufacturing, and human resources are all part of the internal capabilities. The external process concerns consumer demand management and supplier relationships, enabling organizations to increase capabilities through external sources.



**Fig. 1.** Conceptual Model for BI Success (Source: Developed by the Researchers)

## 6 Conclusions

Through a fact-based learning organizational culture and managing change, organizations should focus on strategic alignment of BI and business strategy. The innovative and creative aspects can be determined based on the future development of BI. During the era of technical BI and organization skills, a comprehensive set of soft BI skills should be developed. This study explored how resource-based view theory could be applied to business intelligence. This study provides knowledge on BI and a comprehensive conceptual model indicating that BI success factors may enable organizations to increase their capabilities if they focus on tangible and intangible resources, entrepreneurial resource planning, data management system, and data analytics. These factors will support organizations in developing strategies to adopt technology, manage human resources, and establish processes to manage change, data, and creative culture. The capabilities required as per the resource-based view are strongly dependent on the dynamics of the internal and external environment. The content analysis shows that business intelligence is still viewed primarily as a tool for collecting and analysing data rather than as a catalyst for making better decisions, improving processes, or starting new businesses. Thus, it is recommended that organizations should take BI activities to enhance their capabilities and competencies to gain or sustain competitive advantage. For instance, connecting BI with business strategies will develop a culture based on facts and information and support effective human resource management. This study contributes toward BI importance and discussions from a resource-based perspective, which may help the organizations to bring together strategic management tools and techniques in managing and focusing on internal resources, as well as support in developing the BI approach towards gathering market knowledge in a timely and reliable manner to bring changes and manage them to increase in their capabilities. Future research can test the conceptual model through a quantitative or qualitative study. Moreover, some researchers can study the association between BI and agility through a resource-based approach.

**Acknowledgment.** The author would like to acknowledge the research fund supported by the Muscat College, Sultanate of Oman.

## References

1. Mordor Intelligence. Business Intelligence (BI) Market - Growth, Trends, Covid-19 Impact, and Forecasts (2022–2027) (2021)
2. Al Lawati, M.A.J., Tewari, V.: Business intelligence as a market driver for the personal care brands: case of the Middle East. *Asian J. Adv. Res.* 245–260 (2021)
3. Chatterjee, S., Rana, N.P., Dwivedi, Y.K.: How does business analytics contribute to organizational performance and business value? A resource-based view. *Inf. Technol. People* (2021)
4. El-Adaileh, N.A., Foster, S.: Successful business intelligence implementation: a systematic literature review. *J. Work-Appl. Manage.* **11**, 121–132 (2019)
5. Rao, D.D., Provodnikova, A.: Analysing the role of business analytics adoption on effective entrepreneurship. *Glob. J. Bus. Integr. Secur.* (2021)
6. Alnoukari, M., Hanano, A.: Integration of business intelligence with corporate strategic management. *J. Intell. Stud. Bus.* **7**(2) (2017)
7. Bogdana, P.I., Felicia, A., Delia, B.: The role of business intelligence in business performance management. *Ann. Fac. Econ.* **4**(1), 1025–1029 (2009)
8. Stackowiak, R., Rayman, J., Greenwald, R.: *Oracle Data Warehousing & Business Intelligence SO*. Wiley, Hoboken (2007)
9. Teece, D.J.: Business models and dynamic capabilities. *Long-Range Plann.* **51**(1), 40–49 (2018)
10. Mikalef, P., Pappas, I.O., Krogstie, J., Giannakos, M.: Big data analytics capabilities: a systematic literature review and research agenda. *IseB* **16**(3), 547–578 (2018)
11. Arefin, M.S., Hoque, M.R., Bao, Y.: The impact of business intelligence on organization's effectiveness: an empirical study. *J. Syst. Inf. Technol.* **17**(3), 263–285 (2015)
12. Wixom, B., Watson, H.: The BI-based organization. *Int. J. Bus. Intell. Res. (IJBIR)* **1**(1), 13–28 (2010)
13. Khan, R.A., Quadri, S.: Business intelligence: an integrated approach. *Bus. Intell. J.* **5**(1), 64–70 (2012)
14. Singh, R., Sharma, P., Foropon, C., Belal, H.: The role of big data and predictive analytics in employee retention: a resource-based view. *Int. J. Manpower* **43**, 411–447 (2022)
15. Sarkhani Ganji, H.R., Najafi-Moghadam, A., Sarraf, F.: The impact of management accounting systems on development of intellectual capital dimensions by emphasis on business intelligence in Iran capital market. *Int. J. Financ. Manag. Acc.* **7**(26), 133–143 (2022)
16. Rahimi Holori, B., Ahmadi, F., Khan Mohammadi, M.H., Ranjbar, M.H., Kordlouie, H.: Provide a model of management accounting information system based on business intelligence based on grounded theory. *J. Manage. Acc. Auditing Knowl.* **11**(42), 357–368 (2022)
17. Appelbaum, D., Kogan, A., Vasarhelyi, M., Yan, Z.: Impact of business analytics and enterprise systems on managerial accounting. *Int. J. Account. Inf. Syst.* **25**, 29–44 (2017)
18. Eiden, A., Eickhoff, T., Göbel, J., Apostolov, C., Savarino, P., Dickopf, T.: Data networking for industrial data analysis based on a data backbone system. *Proc. Des. Soc.* **2**, 693–702 (2022)
19. Santos, M., João, E., Canelas, J., Bernardino, J., Pedrosa, I.: The incorporation of business intelligence with enterprise resource planning in SMEs. In: 2021 16th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1–6. IEEE (2021)

20. Becerra-Godinez, J.A., Serralde-Coloapa, J.L., Ulloa-Marquez, M.S., Gordillo-Mejia, A., Acosta-Gonzaga, E.: Identifying the main factors involved in business intelligence implementation in SMEs. *Bull. Elect. Eng. Inform.* **9**(1), 304–310 (2020)
21. Maleki, R., Sabet, E.: Business intelligence analysis in small and medium enterprises. *Int. J. Innov. Mark. Elem.* **2**(1), 1–11 (2022)
22. Saggi, M.K., Jain, S.: A survey towards an integration of big data analytics to big insights for value-creation. *Inf. Process. Manage.* **54**(5), 758–790 (2018)
23. Niu, Y., Ying, L., Yang, J., Bao, M., Sivaparthipan, C.: Organizational business intelligence and decision making using big data analytics. *Inf. Process. Manage.* **58**(6), 102725 (2021)
24. Sun, Z., Huo, Y.: The spectrum of big data analytics. *J. Comput. Inf. Syst.* **61**(2), 154–162 (2021)
25. Huang, Z.-X., Savita, K., Zhong-jie, J.: The business intelligence impact on the financial performance of start-ups. *Inf. Process. Manage.* **59**(1), 102761 (2022)
26. Shamim, S., Yang, Y., Zia, N.U., Shah, M.H.: Big data management capabilities in the hospitality sector: service innovation and customer-generated online quality ratings. *Comput. Hum. Behav.* **121**, 106777 (2021)
27. Bayrak, T.: A review of business analytics: a business enabler or another passing-fad. *Procedia Soc. Behav. Sci.* **195**, 230–239 (2015)
28. Manyika, J., et al.: Big data: the next frontier for innovation, competition, and productivity. McKinsey Global Institute (2011)
29. Paradza, D., Daramola, O.: Business intelligence and business value in organisations: a systematic literature review. *Sustainability* **13**(20), 11382 (2021)
30. Işık, Ö., Jones, M.C., Sidorova, A.: Business intelligence success: the roles of BI capabilities and decision environments. *Inf. Manage.* **50**(1), 13–23 (2013)
31. Beck, J.B., Wiersema, M.F.: Executive decision making: linking dynamic managerial capabilities to the resource portfolio and strategic outcomes. *J. Leadersh. Organ. Stud.* **20**(4), 408–419 (2013)



# Predictive Modeling of Breast Cancer Subtypes Using Machine Learning Algorithms

Ashima Aggarwal<sup>(✉)</sup> and Anurag Sharma

School of Engineering, Design and Automation, GNA University, Phagwara, India  
ashimasagitararius@gmail.com

**Abstract.** In this study, we aimed to classify breast cancer patients into four molecular subtypes: Luminal A, Luminal B, Her-2, and triple negative, using six machine learning techniques: logistic regression (LR), naive Bayes (NB), k-nearest neighbors (KNN), support vector machine (SVM), decision tree (DT), and random forest (RF). We evaluated the performance of each model using several evaluation metrics, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC). The dataset used in this study was obtained in real-time from breast cancer patients, and includes immunohistochemistry (IHC) marker reports. Our results show that all six models achieved high accuracy and AUC scores, indicating their effectiveness in classifying breast cancer patients into molecular subtypes. However, the random forest model outperformed the other models with an AUC score of + 0.95, followed by Logistic Regression with an AUC score of 0.91. These findings demonstrate the potential of machine learning techniques in accurately classifying breast cancer patients into molecular subtypes, which could inform clinical decision-making and personalized treatment strategies.

**Keywords:** Breast cancer · molecular classification · luminal A · luminal B · HER2 · triple negative · machine learning · logistic regression · Naive Bayes · KNN · SVM · decision tree · random forest

## 1 Introduction

Breast cancer is the most commonly diagnosed cancer and the leading cause of cancer death among women worldwide. It is a highly heterogeneous disease with different subtypes, which can be characterized based on molecular markers. The molecular classification of breast cancer into luminal A, luminal B, HER2-enriched, and triple-negative subtypes has important implications for prognosis and treatment.

Machine learning techniques have been widely used for the molecular classification of breast cancer, as they can effectively analyze large amounts of data and identify patterns that are difficult for human experts to detect. In this study, we compare the performance of six machine learning techniques for the molecular classification of breast cancer: logistic regression (LR), naive Bayes (NB), k-nearest neighbors (KNN), support vector machines (SVM), decision tree (DT), and random forest (RF).

The rest of the paper is organized as follows. The related work is represented in Sect. 2. In Sect. 3, we provide a brief overview of the molecular classification of breast cancer and the six machine learning techniques used in this study. In Sect. 4, we describe the dataset used for our experiments. In Sect. 5, we present our experimental results and compare the performance of the six machine learning techniques. Finally, we conclude the paper in Sect. 6.

## 2 Related Work

Breast cancer is a complex and heterogeneous disease that requires personalized treatment strategies. Several studies have compared different machine learning algorithms for breast cancer subtyping.

Perou et al. (2000) were among the first to propose a molecular classification of breast tumors based on gene expression profiles. They identified four intrinsic subtypes: luminal A, luminal B, HER2-enriched, and basal-like. Subsequent studies have refined this classification and identified additional subtypes, such as normal-like and claudin-low (Sørlie et al., 2001; Prat et al., 2013).

Several studies have used machine learning algorithms to classify breast cancer subtypes based on gene expression data. Shi et al. (2005) used a microarray gene expression signature to classify breast tumors into four subtypes: luminal A, luminal B, HER2-enriched, and basal-like. Bhardwaj et al. (2020) compared several machine learning algorithms and found that random forest had the highest accuracy for predicting breast cancer subtypes. Liu et al. (2021) used a deep neural network to accurately classify breast cancer subtypes based on gene expression data.

In addition to gene expression data, histopathological images have also been used to classify breast cancer subtypes. Yamashita et al. (2018) used deep learning algorithms to analyze immunohistochemistry images and identified nuclear features that could predict clinical outcomes in breast cancer patients. Zhang et al. (2018) also used deep learning algorithms to extract features from tumor histopathological images and identified three subtypes: luminal, HER2-enriched, and basal-like. Prat et al. (2012) used IHC markers to classify breast cancer patients into five different subtypes: Luminal A, Luminal B, Her-2 enriched, basal-like, and normal-like. The authors found that the Luminal A subtype was associated with better survival outcomes compared to the other subtypes.

Zhang et al. (2020) used radiomics features and T2\*-weighted imaging to identify triple-negative breast cancer subtypes and develop a preoperative nomogram. Chen et al. (2018) reviewed the use of deep neural networks in breast cancer diagnosis assistance and identified potential applications for radiomics data. Smaili et al. (2021) used an optimized gene expression dataset and compared several algorithms, including K-means clustering, decision tree, random forest, and support vector machine, to identify breast cancer subtypes. Xu et al. (2019) compared several machine learning algorithms and found that support vector machine had the highest accuracy for molecular classification of breast cancer. Overall, machine learning algorithms have shown promise in accurately classifying breast cancer subtypes.

### 3 System Model

Breast cancer is a heterogeneous disease that can be classified into several molecular subtypes based on the expression of specific genes and proteins. The most commonly used classification system is based on the expression of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and cellular proliferation marker(Ki-67). These four biomarkers are used to define four major subtypes: luminal A, luminal B, HER2-enriched, and triple-negative/basal-like.

Machine learning (ML) techniques have shown promise in accurately classifying breast cancer molecular subtypes based on gene expression profiles. In this study, six ML techniques were used: logistic regression (LR), naive Bayes (NB), k-nearest neighbors (KNN), support vector machines (SVM), decision tree (DT), and random forest (RF). These techniques were used to build classification models based on gene expression data and to predict the molecular subtypes of breast cancer samples. The performance of each ML technique was evaluated using several metrics such as accuracy, precision, recall, F1-score, confusion matrix, and area under the curve (AUC).

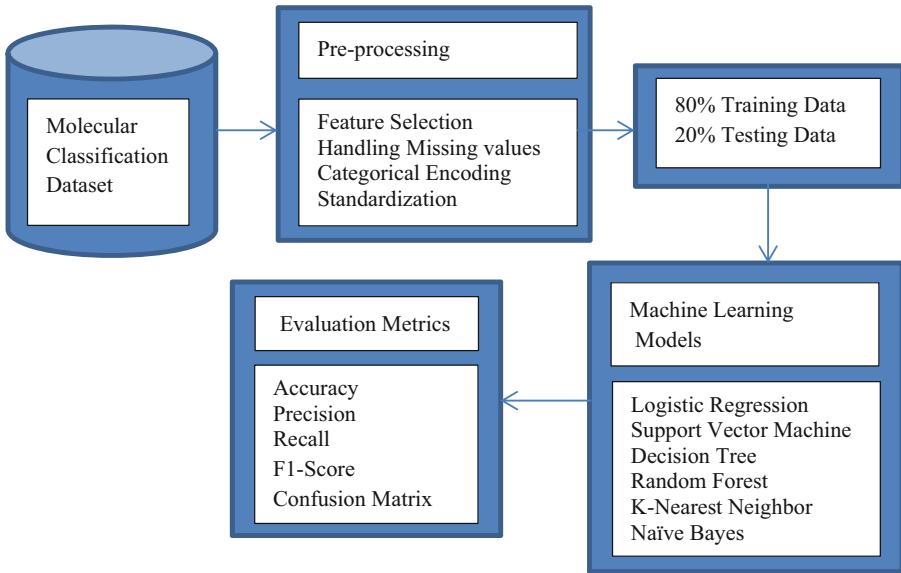


Fig. 1. Flow diagram of the proposed method

#### Here’s a Brief Overview of the Six Machine Learning Techniques used in this Study:

**Logistic Regression (LR):** A popular classification algorithm used for binary and multi-class classification problems. It models the probability of the target variable using a linear combination of input features.

**Naive Bayes (NB):** A probabilistic algorithm that uses Bayes' theorem to predict the probability of each class based on the input features. It assumes that the features are independent of each other, hence the name "naive."

**K-Nearest Neighbors (KNN):** A non-parametric algorithm that predicts the class of a data point based on the class labels of its k-nearest neighbors in the training data.

**Support Vector Machine (SVM):** A powerful algorithm for both classification and regression problems. It finds the optimal hyperplane that separates the different classes by maximizing the margin between them.

**Decision Tree (DT):** A tree-based algorithm that recursively partitions the data into subsets based on the most informative features, creating a tree-like model of decisions and their possible consequences.

**Random Forest (RF):** A type of ensemble learning method that combines multiple decision trees to create a more accurate and robust model. It randomly selects subsets of features and samples from the data to build each tree.

## 4 Dataset

The dataset used in this study is a real-time dataset obtained from breast cancer patients, and it includes immunohistochemistry (IHC) marker reports. The dataset consists of 146 samples, with each sample containing information on four IHC markers: ER, PR, Ki-67, and Her-2 Neu.

The ER (estrogen receptor) and PR (progesterone receptor) markers are used to determine whether the breast cancer cells are hormone receptor positive or negative. The Ki-67 marker is a marker of cellular proliferation and is used to predict the growth rate of the cancer cells. The Her-2 Neu marker is used to identify whether the cancer cells overexpress the Her-2 protein, which is associated with aggressive breast cancer.

In addition to the IHC markers, the dataset also includes the molecular classification of each sample, which was used as the ground truth for the machine learning models. The samples were classified into four molecular subtypes: Luminal A, Luminal B, Her-2, and triple negative.

## 5 Experimental Results

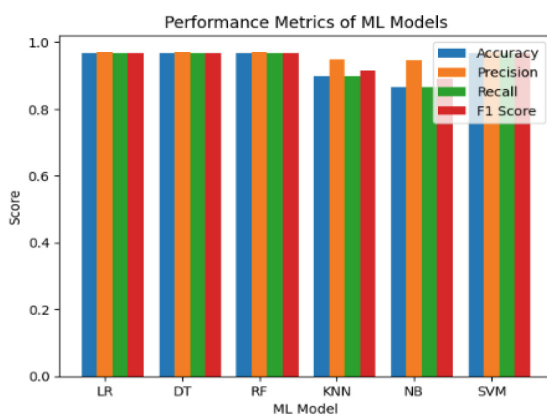
The accuracy score measures the percentage of correctly classified cases and accuracy of different machine learning models is shown in Fig. 1. And precision measures the proportion of true positive classifications out of all positive classifications. Recall measures the proportion of true positive classifications out of all actual positive cases, and F1 score is the harmonic mean of precision and recall.

The Table 1 shows the performance metrics of six machine learning models used for the molecular classification of breast cancer. The metrics evaluated for each model include accuracy, precision, recall, and F1 score.



**Table 1.** Performance Metrics of Machine Learning Models for Molecular Classification of Breast Cancer

ML model/Metrics	Accuracy Score	Precision Score	Recall Score	F1 Score
LR	0.97	0.97	0.97	0.97
DT	0.97	0.97	0.97	0.97
RF	0.97	0.97	0.97	0.97
KNN	0.9	0.95	0.9	0.91
NB	0.87	0.95	0.87	0.89
SVM	0.97	0.97	0.97	0.97

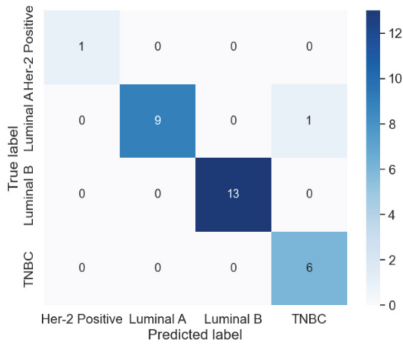
**Fig. 2.** Evaluation Metrics of Machine Learning Models on Molecular Classification Dataset

Based on the results in the table, it can be seen that all models achieved high accuracy scores ranging from 0.866 to 0.967. The models also showed high precision, recall, and F1 scores, indicating their effectiveness in classifying breast cancer patients into molecular subtypes. Performance of Machine Learning Models on Molecular Classification Dataset is shown in Fig. 2. The random forest, logistic regression, decision tree and SVM models achieved the highest scores across all metrics, with all models achieving an accuracy score of 0.9667 and precision, recall, and F1 scores of 0.9714, 0.9667, and 0.9671, respectively. The KNN model achieved slightly lower scores but still showed high performance, with an accuracy score of 0.9 and precision, recall, and F1 scores of 0.9492, 0.9, and 0.9143, respectively.

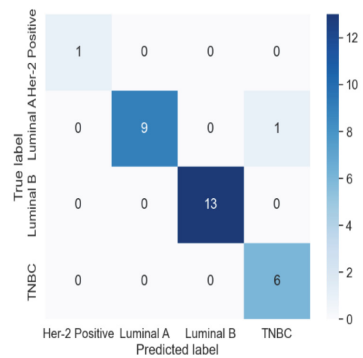
Overall, these results demonstrate the potential of machine learning techniques in accurately classifying breast cancer patients into molecular subtypes, which could inform clinical decision-making and personalized treatment strategies.

**Confusion Matrix:** A confusion matrix is a table used to evaluate the performance of a classification model by comparing the predicted and true labels for a set of data. In a

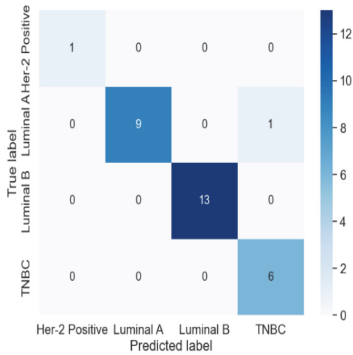
binary classification problem, the confusion matrix has four components: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).



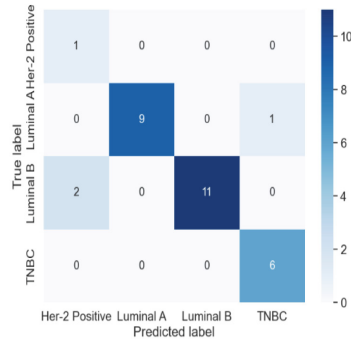
Confusion matrix for Logistic Regression model



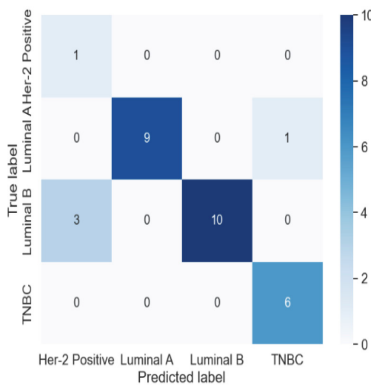
Confusion matrix for Decision Tree model



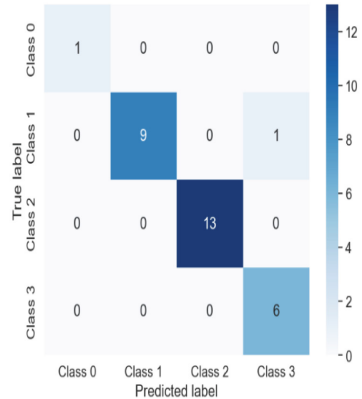
Confusion matrix for Random Forest model



Confusion Matrix of KNN model



Confusion matrix for Naive Bayes model



Confusion matrix for SVM model

**Fig. 3.** Confusion Matrix of all machine learning models

In a multi-class classification problem, such as the one represented by the given confusion matrix for logistic regression, the confusion matrix is a square matrix with the number of rows and columns equal to the number of classes. The diagonal of the matrix represents the number of correctly classified samples for each class (true positives), while the off-diagonal elements represent the misclassified samples. Confusion matrix of all machine learning models is shown in Fig. 3. Here is the explanation of one Confusion matrix. And same is the way to understand all confusion matrix.

**In Logistic Regression Confusion Matrix,** For class Her-2 Positive, there is one true positive, meaning that one instance of this class was correctly classified as such. For class Luminal A, there are nine true positives and one false positive, meaning that nine instances of this class were correctly classified as Luminal A, but one instance of another class was incorrectly classified as Luminal A. For class Luminal B, there are thirteen true positives, meaning that all thirteen instances of this class were correctly classified. For class TNBC, there are six true positives, meaning that all six instances of this class were correctly classified.

The Graphical user interface (GUI) for Molecular Classification of Breast Cancer is shown in Fig. 4.

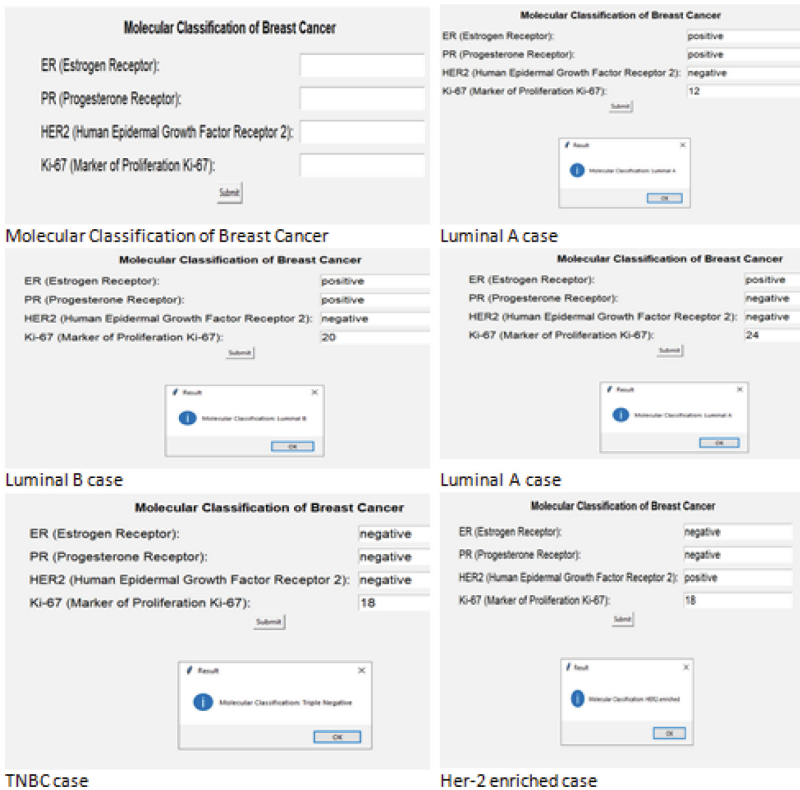


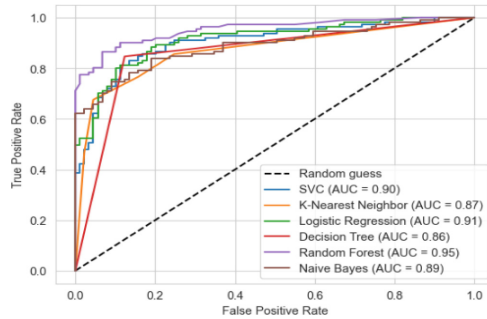
Fig. 4. GUI for Molecular Classification of Breast Cancer

## 6 Conclusion

In conclusion, the molecular classification of breast cancer into luminal A, luminal B, HER-2, and triple negative is an important step towards personalized treatment of breast cancer patients. In this study, we used six machine learning techniques, namely logistic regression, Naive Bayes, KNN, SVM, decision tree, and random forest, to classify breast cancer patients based on their molecular subtypes.

The AUC score represents the area under the Receiver Operating Characteristic (ROC) curve, which is a graphical representation of the performance of a binary classifier as the discrimination threshold is varied. AUC score of all machine learning models can be seen in Fig. 5. An AUC score of 0.91 indicates that the model is able to distinguish between positive and negative instances with high accuracy. However, it's important to consider the context of the problem and the consequences of misclassifications before making a final conclusion.

Based on the evaluation metrics provided, all the models have achieved a high accuracy score of at least 0.8667. Additionally, the precision, recall, and F1 scores are also high and consistent across all the models. This indicates that the models are able to classify the instances with high accuracy and are reliable. Looking at the confusion matrices, it is also evident that there are very few misclassifications across all the models. However, the KNN and Naive Bayes models have misclassified a few instances, which are reflected in their lower accuracy score compared to the other models. The AUC scores provide an additional metric to evaluate the models' performance, and the Random Forest model achieved the highest AUC score of 0.95, followed closely by the Logistic Regression and SVM models with an AUC score of 0.91.



**Fig. 5.** AUC score of machine learning models

In conclusion, all the models perform well on the given dataset, but the Random Forest model has the highest AUC score and may be the best model to choose for this classification problem.

Overall, these findings suggest that machine learning techniques can be valuable tools in the molecular classification of breast cancer and can aid in the development of personalized treatment plans for breast cancer patients. Further research is needed to explore the potential of these techniques in clinical practice and to validate their findings.

## References

- Breast Cancer Molecular Subtypes, cancercenter.com, last accessed on (18 Feb 2023)
- Liu, Y., Lin, Y., Yang, X., Dong, Y., Li, X., Liu, J.: Accurate classification of breast cancer subtypes using deep neural network and machine learning algorithms. *BMC Med. Inform. Decis. Mak.* **21**(1), 1–10 (2021)
- Smaili, H., Touns, N., Zouaki, H., et al.: Identification of breast cancer molecular subtypes using machine learning techniques based on optimized gene expression data. *Cancer Med.* **10**(1), 341–352 (2021)
- Bhardwaj, M., Srivastava, S., Singh, V., et al.: Breast cancer molecular subtypes prediction by machine learning methods using transcriptomic data. *BMC Bioinformatics*. (Suppl 13), 425 (2020–2021)
- Zhang, Y., Zhao, Y., Huang, J., Yu, L.: Identification of triple-negative breast cancer subtypes and preoperative nomogram using a machine learning approach based on radiomics and T2\*-weighted imaging. *Front. Oncol.* **10**, 1057 (2020)
- Wang, Z., Yang, Y., Yang, D., et al.: Deep learning models for breast cancer classification using gene expression profiles. *BMC Med Genomics* **12**(Suppl 5), 108 (2019)
- Hon, J.D., Singh, B., Sahin, A., et al.: Breast Cancer Molecular Subtypes: From TNBC to QNBC. *Am J Cancer Res* (2019)
- Liu, C., Shi, W., Zhang, L., Wu, L., Shen, Y.: A support vector machine based classifier for breast cancer subtype prediction using RNA-Seq data. *IEEE Access* **7**, 116364–116372 (2019)
- Xu, J., Liu, J., Xue, F., Li, J.: Application of support vector machine in molecular classification of breast cancer based on gene expression data. *Journal of Biomedical Informatics* **91**, 103136 (2019)
- Zhang, Y., Xu, Y., Wang, W., Liu, Q.: Application of decision tree algorithm in the classification of breast cancer in young women. *Journal of Healthcare Engineering* **2019**, 1–10 (2019)
- Chen, H., Li, Q., Liu, X., Wu, X., Shi, H.: Deep neural network model for breast cancer diagnosis assistance: A review and a vision. *Artif. Intell. Med.* **89**, 1–9 (2018)
- Dillon, D.A., Hassell, L.A.: Molecular subtypes of breast cancer: a review for breast radiologists. *Journal of breast imaging* **1**(2), 87–96 (2018). <https://doi.org/10.1093/jbi/wby012>
- Yamashita, R., Nakano, K., Ueno, H., Mitsuyama, S., Nishimura, R.: Deep learning of immunohistochemistry images reveals nuclear features for clinical outcome prediction of patients with breast cancer. *Sci. Rep.* **8**(1), 1–9 (2018)
- Zhang, Y., Zhu, W., Yang, L., Wu, J., Luo, X.: Breast cancer molecular subtyping using deep features learned from tumor histopathological images and its association with prognosis. *Journal of Clinical Oncology* **36**(15\_suppl), e12519–e12519 (2018)
- Jhang, X.: Molecular classification of breast cancer: relevance and challenges. *The J. Pathol. Translat. Med.* **51**(1), 1–12 (2017)
- Ades, F., Zardavas, D., Bozovic-Spasojevic, I., et al.: Molecular classification of breast cancer: where do we stand? *Lancet Oncol.* **15**(3), e216–e226 (2014). [https://doi.org/10.1016/S1470-2045\(13\)70540-1](https://doi.org/10.1016/S1470-2045(13)70540-1)
- Prat, A., Parker, J.S., Perou, C.M.: Molecular classification of breast cancer. *Oncologist* **18**(4), 326–335 (2013)
- Prat, A., Perou, C.M., Mamounas, E.P.: Molecular classification of breast cancer: what the clinician needs to know. *The surgeon* **10**(6), 336–342 (2012)
- Shi, Y., Huang, H.C., Zhou, H., et al.: Breast cancer classification using machine learning algorithms with a microarray gene expression signature. *Cancer Inform.* **1**, 108–116 (2005)
- Sørli, T., et al.: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci.* **98**(19), 10869–10874 (2001)
- Perou, C.M., et al.: Molecular portraits of human breast tumours. *Nature* **406**(6797), 747–752 (2000)



# A Systematic Study on Unimodal and Multimodal Human Computer Interface for Emotion Recognition

Akram Ahmad<sup>1</sup>, Vaishali Singh<sup>2</sup>, and Kamal Upreti<sup>3</sup>(✉)

<sup>1</sup> Department of Computer Science, Maharishi University of Information Technology, Lucknow, India

<sup>2</sup> Department of Computer Science, Maharishi University of Information Technology, Lucknow, India

<sup>3</sup> Department of Computer Science, CHRIST (Deemed to Be University), Delhi NCR, Ghaziabad, India

kamalupreti1989@gmail.com

**Abstract.** A systematic study for human-computer interface (HCI) for emotion recognition is presented in this paper, with a focus on various methods used to identify and interpret human emotions. It delves into various methods used to identify and interpret human emotions and highlights the limitations of unimodal HCI for emotion recognition systems. The paper emphasizes the benefits of multimodal HCI and how combining different types of data can lead to more accurate results. Additionally, it highlights the importance of using multiple modalities for emotion recognition. The study has significant implications for mental health assessments and interventions as it offers insights into the latest techniques and advancements in emotion recognition. Future research can use these insights to improve the accuracy of emotion recognition systems, ultimately leading to better mental health assessments and interventions. Overall, the paper provides a valuable contribution to the field of HCI and emotion recognition, and it underscores the importance of taking a multimodal approach for this critical area of research.

**Keywords:** Human Computer Interface · Emotion Recognition · Unimodal · Multimodal · Machine Learning

## 1 Introduction

Emotions play a crucial role in interpersonal relationships, perception, information, understanding, and other facets of existence [1]. As a result, the ability to understand emotions has become essential for human functioning, especially in social settings where emotions are necessary for communication. Automated emotion identification systems are increasingly in demand, making emotion recognition of a crucial field of research in human-computer interfaces [2]. Researchers have used voice, texts, facial expressions, and EEG-based brain waves to study emotion recognition. Emotion recognition technology can enhance user interactions by providing agreeable or intuitive user experiences.

The market value for emotion identification and recognition was USD 19.87 million in 2020. Utilizing emotional signals from multiple modalities can help forecast an individual's emotional state since a single modality struggles to assess emotions accurately. One can't determine if somebody is feeling something just by glancing at something or something happening in front of us. Due to the challenge of developing emotion-specific characteristics, emotion detection is a multimodal task. The fact that various individuals display their sentiments in various ways, non-linear encounters among various mediums, and multi-dimensional information are the causes of this intricacy. Through the detection of intricate non-linear characteristic correlations in multimedia information, deep networks and machine learning have shown to be successful in addressing these constraints [3].

Traditional feature engineering and machine learning may not extract complex trends from multivariate time series data, requiring dimensionality reduction techniques. DL methods, such as convolutional neural networks, recurrent neural networks, and autoencoders, provide high-level data abstraction and improve algorithms in emotion detection. These methods extract unique features from high-level data representations, enhancing accuracy. Although most emotion acquisition systems use 2D non-immersive visual aids such as pictures or videos to elicit emotional states, immersive virtual reality (VR) is becoming increasingly popular in emotion recognition. Researchers can replicate experiences in controlled laboratory settings using VR, while also creating a sense of presence and involvement. By combining implicit measures from VR with machine learning methods, researchers can explore alternative approaches that could have a significant impact on various research areas, presenting new opportunities for researchers [4, 5].

Several methods are employed to detect and understand human emotions, including analyzing facial expressions, speech patterns, and physiological signals [6]. However, nonlinear patterns in multivariate time-series data can present challenges for traditional attribute engineering and machine learning approaches, requiring the use of deep learning (DL) techniques like convolutional neural networks (CNN), recurrent neural networks (RNN), and autoencoders. The emergence of immersive virtual reality (VR) and sensor data presents new opportunities for emotion recognition systems to improve accuracy by simulating laboratory conditions and analyzing social behaviors, respectively, as noted in [7]. As a result, the future of emotion recognition looks bright, with the development of new technologies that promise to enhance accuracy and provide deeper insights into the complexities of human emotions.

## 2 Overview of HCI

HCI stands for human computer interface that is dedicated to study, design, develop and evaluate the interaction of machines with human and interprets human emotional state. It aims to make computing more accessible, usable, and efficient by taking into account users' needs, preferences, and limitations. HCI principles are drawn from a variety of disciplines, including computer science, psychology, sociology, anthropology, and design. The goal of HCI is to design user interfaces that are simple, efficient, and enjoyable to use. Understanding the user's goals and tasks, designing interfaces to support those goals and tasks, and evaluating the usability of those interfaces are all part of this process.

Interface design, usability testing, cognitive psychology, human factors, and social computing are all topics covered in HCI research [8]. Voice and pictorial identification are not the only applications of multimodal technological capabilities, as it has the potential to transform the entire world. Various industries have already implemented several leading multimodal interaction technologies, incorporating voice identification, lip reading, audio conversion, and language understanding, and gesture interaction technology. Gesture interaction technology has become a vital method of HCI, alongside traditional input devices like keyboards, mice, and touch screens. By recording the motions of person's limbs and arms, gesture interaction technology transforms commands into a linguistic structure that computers can understand. In the field of intelligent hardware, microphone arrays are the mainstream method for processing signals and reducing noise, but there are still significant challenges in speech recognition when faced with complex and noisy environments [9]. The upcoming breakthrough in HCI technology might not have the exact similar wide-ranging effects as the advent of graphical user interfaces and touchscreen technologies. Instead, it may leverage data-driven intelligence to unlock the potential for HCI. Artificial intelligence is rapidly advancing, and the study of human-machine interaction is promoting automated voice identification technique and innovative gestural interaction technologies. Overall, multimodal interaction technology has the potential to change the entire world, and researchers and pioneers are working tirelessly to realize this potential. By leveraging the power of data-driven intelligence and advancing artificial intelligence, we can create a new era of HCI that improves the way humans interact with technology [10].

Human-Computer Interaction (HCI) is an interdisciplinary field of research that deals with the interface between individuals and machines. The term "HCI" was primary given in 1975, and the proficient term for this field of study appeared in 1980 [11]. Since then, the popularity of the concept of HCI has led to an increase in research in this area, with deep learning technology further accelerating the process [12]. HCI can be simply defined as the way in which people and machines interact with each other. In recent years, the focus of HCI research has moved from command-based interaction to emotional interaction. However, there are still many challenges that need to be addressed in the progress of this communication. Some of the key challenges facing HCI research in this area include developing effective speech recognition systems that can identify background noise and actual communicating resonances, improving machine learning algorithms to better comprehend human language, accurately identifying conscious and unconscious gestures, etc. [13]. Despite these challenges, the focus of HCI research has shifted from humans adapting to machines to machines alter as per the humans. Dialog robot commodities relying on artificial intelligence technique are slowly becoming mainstream [14]. HCI deals with the communication among humans and machines. The growth of deep learning technology has accelerated research in this area, and the focus has transferred from command-based interaction to emotional interaction. The virtual world poses unique challenges for HCI, and there is still much work to be done in developing effective voice and gesture-based interaction methods. The future of HCI research looks promising, and we can expect continued advances in this area in the years to come.



### 3 Unimodal HCI

Emotion detection is a technology that uses various techniques like as biosensors, deep learning, face identification, speech identification, voice identification, and pattern identification to identify and recognize human emotions. Initial emotion comprehension relying on a self-assessment set of questions, voice pitch, or facial appearance, on the other hand, can be ambiguous and partial. Furthermore, people can intentionally conceal or analyse their real feelings, attempting to make human-based emotion recognition susceptible to mistakes. As a consequence, physiological signal assessments including heart rate, electromyogram, and galvanic skin reaction have been employed more frequently in emotion detection methods [5]. The two types of emotion detection techniques are single and multimodal. Single modality emotion detection involves using a single modality, such as facial expressions or speech, but it may have limitations in specific contexts. Multimodal emotion detection techniques, which use a variety of input data and identifies the emotion more accurately and tried to overcome the limitations of single modality.

#### 3.1 Textual Based Unimodal HCI for Emotion Recognition

Textual emotion-mining techniques have numerous practical applications, such as determining customer satisfaction levels, aiding in the selection of e-learning materials, suggesting commodities based on user emotions, and even forecasting mental health illnesses. However, existing sentiment examination surveys do not adequately address the deep connection between opinion mining and emotion mining. Consequently, a fresh perspective on sentiment analysis literature is necessary, emphasizing emotion mining. Yadollahi et al. [15] provides state-of-the-art sentiment assesment methods, including (1) a sentiment analysis taxonomy; (2) a comprehensive survey of polarity cataloguing techniques and resources, with a focus on emotion mining; (3) a comprehensive survey of emotion theories and emotion-mining research; and (4) several helpful resources, like lexicons and databases.

In their study, Binali et al. [16] established emotion theories that serve as a foundation for emotion models, which have been utilized to develop computational techniques for emotion identification. In response, they proposed a hybrid architecture for emotion identification. Canales et al. [17] conducted a literature review on lexical and machine learning approaches and categorized them based on the emotional model and approach employed. Kang et al. [18] utilized contextual information to predict emotions for both word and document text, using Bayesian models. Perikos et al. [19] designed an ensemble classifier schema for sentiment assessment, integrating knowledge-based and statistical machine learning approaches to optimize their strengths and decrease their limitations. Finally, Nahin et al. [20] accumulated consumer passwords and emotional responses in the presence and absence of particular messages in order to instruct and examine classification techniques. These researches show the prominence of considering emotion theories and selecting appropriate approaches when designing emotion detection systems. Md. Taufeeq Uddin et al. [21] proposed a Wrapper Random Forest-based characteristic selection approach for speech emotion recognition. They used Gabor filters to extract voice parameters and applied feature selection on a broad range of features.

The Random Forest algorithm was then applied to generate importance scores for the voice sequence characteristics, and a multi-class SVM was used to train the chosen features on the CK database. The results showed that the method of feature selection could increase the recognition of emotions like sadness, neutral, and anger. Using a Cauchy Bayes Classifier, NicuSebe et al. [22] identified spoken emotions and showed them as video sequences. Martin Vondra et al. [23] proposed a method for recognizing emotions from text using a knowledge-based artificial neural network (KBANN).

### 3.2 Video Based Unimodal HCI for Emotion Recognition

The computer vision-based emotion recognition method detects emotions from visual media, but current models lack inter-feature interaction and rely on basic concatenation for feature fusion, which can result in misunderstanding and confusion. Hence, there is a need for improved models that can integrate both facial and contextual features for better accuracy in detecting emotions [24].

Bounyong et al. [25] conducted study on remotely with 100 Japanese students, where video constituents were used as stimuli to induce epistemic emotions, while facial video and HRV were documented. To gather the information, the authors modified the technique used in earlier investigations. The findings demonstrated that the algorithms developed using the merged characteristics were capable of reliably estimating the six epistemic emotions. Nasir et al. [26] employed geometric distance signals in a different investigation to identify six fundamental emotions. From all the triangles generated by the three prominent facial features that were chosen, the writers looked at four centre locations. Six alternative geometric distance signatures were recovered in order to express various geometric representations of an emotional image sequence. By using every distance signature individually as input characteristic sets into an MLP classification model, the researchers separately assessed the exclusionary value each one's. The outcomes demonstrated that the given approach attained high precision in recognizing six basic emotions. To accurately recognise emotions from lengthy video clips, Zhou et al. [27] proposed a hybrid feature weighting network. To enhance the effectiveness of emotion detection, the investigators took advantage of the complementing data between facial and environment aspects. A deep fusion (DF) block, a hierarchical attention (HAE) network, and a dual-branch encoding (DBE) network make up the suggested framework. The DBE network's facial and context encoding blocks produce the corresponding shallow characteristics. With a cross-channel attention operation, the HAE network investigates and captures the complementarity between face expression characteristics and their settings using the cross-attention (CA) block. The deep fusion block combines the features from both the face and context encoding blocks to achieve accurate emotion recognition. The studies discussed above propose various methods to recognize emotions from facial and contextual features. These studies provide important insights into the development of accurate and efficient emotion recognition systems. By combining features from videos and HRV, exploring geometric distance signatures, and exploiting the additional data between face and context characteristics, these studies achieve high accuracy in recognizing different emotions. These outcomes could have substantial implications for the development of emotion recognition systems in various

domains, including healthcare, education, and entertainment. Below Table 1 and Fig. 1 represents comparative analysis of some unimodal HCI for emotion recognition.

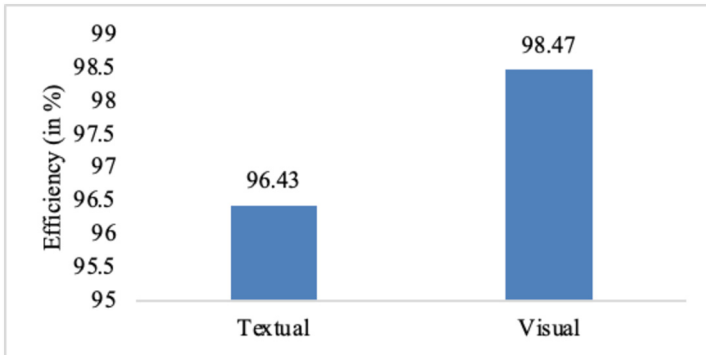
**Table 1.** Unimodal HCI for Emotion Recognition

Authors	Method Used	Emotions	Key Features
Yadollahi et al. [15]	Existing emotion identification approaches	Text	The field of emotion mining on Twitter offers various resources, such as lexicons, datasets, and mining methods, for both English and other languages. By organizing and categorizing emotion mining techniques, researchers can gain a better understanding of this field
Binali et al. [16]	Keyword and learning-based approaches	Text	Computational approaches for detecting emotions from text involve analyzing the words and phrases used to determine the emotional state of the writer. Combining both semantic and syntactic information can enhance emotion detection accuracy
Canales and Martinez-Barco [17]	Lexicon with machine learning	Text	The article discusses various techniques used for emotion mining on Twitter, including Lexical-based, Supervised-learning-based, and Unsupervised-learning-based techniques. It compares the advantages and disadvantages of lexical-based and machine learning-based techniques and highlights the limitations of current systems
Md. Taufeeq Uddin, et.al [21]	Wrapper Random Forest-based feature selection approach	Text	In a study on speech emotion recognition performance, Gabor filters were utilized to extract voice parameters and reduce the wide feature set through feature selection

*(continued)*

**Table 1.** (continued)

Authors	Method Used	Emotions	Key Features
Bounyong et al. [25]	Monitoring epistemic emotions in e-learning	Video	Created models to predict six emotions (Boredom, Confusion, Curiosity, Frustration, Interest, and Surprise) by combining features extracted from videos and Heart Rate Variability (HRV)
Nasir et al. [26]	Triangulation mechanism for geometric feature generation	Video	They calculated the coordinates of the four centroid points for every triangle formed by selecting a triplet of salient facial points
Zhou et al. [27]	Hybrid network	Video	The proposed approach utilized both facial and contextual features and exploits their complementary information to acquire highly precise emotion identification from large-scale video clips. Specifically, we have designed a cross-attention mechanism and a hybrid characteristic weighting network that work together to effectively combine the two types of features and enhance the overall recognition performance

**Fig. 1.** Comparison of accuracy for Unimodal HCI

## 4 Multimodal HCI for Emotion Recognition

The concept of multimodal emotion recognition involves replicating human abilities to predict emotions by combining biological and behavioural features. Unlike unimodal emotion identification, this approach utilizes multiple data sources to achieve more precise recognition of emotions. Nonetheless, the accuracy of multimodal techniques differs significantly, underscoring the necessity for more resilient approaches. To achieve this, it is crucial to develop effective methods that can efficiently process and integrate data from various modalities, thereby enhancing the accuracy of emotion recognition.

Song and Kim [28] developed a method to identify hidden emotions using a convolutional neural network to analyze video and electroencephalogram data, enhancing emotion detection technology. Povolny et al. [29] proposed a multimodal system integrating voice and video characteristics to improve emotion identification accuracy, while Lu et al. [30] identified sixteen eye movement patterns to recognize emotions using EEG signals. Combining multiple modalities such as video, audio, and EEG signals can improve emotion detection technology and lead to better mental health assessments and interventions. Cimtay et al. [31] and Poria et al. [32] explored the use of multiple modalities for emotion recognition. Cimtay et al. used GSR, facial expressions, and EEG to identify arousal state and genuine emotional states in real-time. Poria et al. combined voice, text, and facial expressions, using a temporal deep CNN and MKL to achieve accurate emotion identification. Both studies highlight the importance of combining various types of data to improve emotion recognition technology. Below Table 2 and Fig. 2 represents comparative analysis of some multimodal HCI for emotion recognition.

**Table 2.** Multi-modal HCI for Emotion Recognition

Authors	Method	Type	Key Features
Song and kim et al. [28]	CNN	Visual and EEG	This refers to instances where a person's facial expression appears neutral, but their underlying emotions are revealed through their biological signals
Povolny et al. [29]	CNN	Visual and Textual	A convolutional neural network was used to enhance the baseline data in video by detecting facial landmarks. The CNN integrates geometrical and appearance information in its activations to locate these landmarks

*(continued)*

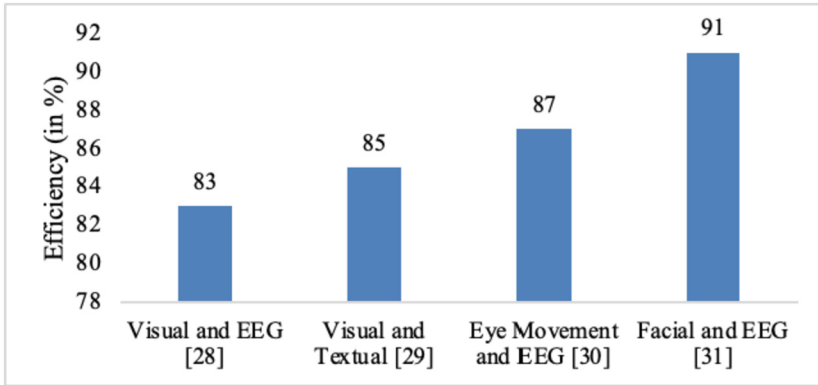
**Table 2.** (continued)

Authors	Method	Type	Key Features
Lu et al. [30]	modality fusion approaches	Eye Movement and EEG	It has been discovered that there are specific patterns of eye movements associated with negative, positive, and neutral emotions. Sixteen of these patterns have been identified. In addition, various modality fusion techniques have shown that both EEG and eye movements could be utilised together to improve the accuracy of emotion recognition. This suggests that combining information from multiple sources can enhance our understanding and detection of emotions
Cimtay et al. [31]	Feature fusion	Facial and EEG	The proposed approach for sentiment identification involves using multiple modalities such as galvanic skin response, facial expressions, and electroencephalogram. The EEG and GSR data are combined through feature fusion to identify the level of provocation

## 5 Current Limitations and Future Research Directions

In the current state-of-the-art, there are various obstacles and chances that have been recognised for the progress of an emotion recognition framework. These challenges and opportunities must be addressed to create a robust framework for emotion recognition [33].

**Experimental Design:** Experimental setups for emotion evocation can be either done in a controlled laboratory environment or naturally obtained in daily-life settings. However, daily-life scenarios present additional challenges, such as difficulty in emotion



**Fig. 2.** Compression of accuracy for Multimodal HCI for Emotion Recognition

awareness, ground-truth data annotation, and a decrement in signal-to-noise ratio because of unconstrained subject movements. This often leads to lower prediction scores in daily-life algorithms. Furthermore, several variables can significantly impact subject response. For instance, the presence of an audience may affect the subject's reaction differently than when alone. Therefore, these variables need to be considered when designing an experimental setup for emotion elicitation, especially in daily-life scenarios.

**EMAs:** Self-annotated reports ought to be straightforward, expeditious, and aimed at achieving specific goals, while also incorporating a system of rewards. Research has shown that the presence of rewards helps to sustain the motivation of participants and increases the likelihood of producing accurate ground truth data.

**Emotion Dimensions:** Current research primarily emphasizes on recognizing emotions in binary form i.e., either yes or no. While accuracy in classifying arousal has been high, there remains a lack of reasonable performance in valence classification. Thus, future research could explore the creation of novel metrics and dimensions for assessing and classifying complex emotions [34, 35].

**Elicitation Material:** Short films are a cost-effective and highly adaptable means of eliciting strong emotional responses in individuals. By focusing on a single emotion, these films can effectively trigger intense emotional reactions. However, it is important to validate these films to ensure their reliability and effectiveness as elicitation materials.

**Sensor Modalities:** Several studies have found that recognition rates improve with an increase in the number of data modalities used, including EEG, ECG, etc. Additionally, there is a lack of publicly accessible databases for emotion identification that incorporate all feasible modalities in real-world, unconstrained daily-living situations.

## 6 Conclusion

Affective computing concepts and studies have grown over the past 13 years. With a better comprehension of emotion hypothesis, the formation of precise machine learning (ML) algorithms, and the ubiquity, speed, and pervasiveness of wearable technology,

these novel technological advances have entered our everyday lives, enhancing our standard of living and making it possible to collect significant proportion of information for the formation of sophisticated ML models for trustworthy emotion detection algorithms. Affective computing has adopted a range of approaches, creating novel possibilities for research, difficulties, and concerns that have increased the awareness of feelings. The effectiveness of the emotion identification procedure depends heavily on the techniques for feature categorization and emotion acquisition. We have given a thorough overview of publically accessible databases for multimodal emotion detection along with a dependable list of emotion collection techniques that concentrate on sensor technologies and virtual reality. Yet, given the lack of regulations to deal with privacy concerns and the discriminatory nature of ML/DL algorithms, we discovered a significant research void. To prevent misclassifying emotion features, which could produce unreliable outcomes, more study must be done. It would be best to create a multimodal database that takes into account circumstances from the actual world. To develop extremely effective emotion identification technologies, more research into optimisation strategies is advised. These advancements have not only made our lives better, but they have also created new study possibilities, difficulties, and issues that have increased our understanding of emotions. Yet, in order to guarantee correct emotion identification and the moral use of affective computing technology, it is essential to overcome the gaps in research and problems in the area, like privacy concerns and bias ML/DL models.

## References

1. Baltrušaitis, T., Ahuja, C., Morency, L.-P.: Multimodal machine learning: a survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**, 423–443 (2019). <https://doi.org/10.1109/TPAMI.2018.2798607>
2. Alswaidan, N., Menai, M.E.B.: A survey of state-of-the-art approaches for emotion recognition in text. *Knowl. Inf. Syst.* **62**, 2937–2987 (2020). <https://doi.org/10.1007/s10115-020-01449-0>
3. Zhang, H.: Expression-EEG Based collaborative multimodal emotion recognition using deep AutoEncoder. *IEEE Access* **8**, 164130–164143 (2020). <https://doi.org/10.1109/ACCESS.2020.3021994>
4. Upreti, K., et al.: IoT-Based Control System to Measure, Analyze, and Track Basic Vital Indicators in Patient Healthcare Monitoring System. In: *Advanced Communication and Intelligent Systems: First International Conference, ICACIS 2022, Virtual Event, October 20–21, 2022, Revised Selected Papers*, pp. 715–725. Springer Nature Switzerland, Cham (2023, February)
5. Tang, H., Liu, W., Zheng, W.L., Lu, B.L.: Multimodal emotion recognition using deep neural networks. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)* 10637 LNCS, 811–819 (2017). [https://doi.org/10.1007/978-3-319-70093-9\\_86](https://doi.org/10.1007/978-3-319-70093-9_86)
6. Marín-Morales, J., Higuera-Trujillo, J.L., Greco, A., et al.: Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. *Sci. Rep.* **8**, 13657 (2018). <https://doi.org/10.1038/s41598-018-32063-4>
7. Long, F., Zhao, S., Wei, X., et al.: Positive and negative emotion classification based on multi-channel. *Front. Behav. Neurosci.* **15**, 1–11 (2021). <https://doi.org/10.3389/fnbeh.2021.720451>
8. Jarosz, M., Nawrocki, P., Śnieżyński, B., Indurkha, B.: Multi-platform intelligent system for multimodal human-computer interaction. *Comput. Inform.* **40**, 83–103 (2021)



9. Ince, G., Yorganci, R., Özkul, A., et al.: An audiovisual interface-based drumming system for multimodal human–robot interaction. *J. Multimodal User Interf.* **15** (2020). <https://doi.org/10.1007/s12193-020-00352-w>
10. Lai, H., Chen, H., Wu, S.: Different contextual window sizes based RNNs for multimodal emotion detection in interactive conversations. *IEEE Access* **8**, 119516–119526 (2020). <https://doi.org/10.1109/ACCESS.2020.3005664>
11. Nayak, S., Nagesh, B., Routray, A., Sarma, M.: A Human–Computer Interaction framework for emotion recognition through time-series thermal video sequences. *Comput. Electr. Eng.* **93**, 107280 (2021). <https://doi.org/10.1016/j.compeleceng.2021.107280>
12. Yuan, J., Feng, Z., Dong, D., et al.: Research on multimodal perceptual navigational virtual and real fusion intelligent experiment equipment and algorithm. *IEEE Access* **8**, 43375–43390 (2020). <https://doi.org/10.1109/ACCESS.2020.2978089>
13. Iio, T., Yoshikawa, Y., Chiba, M., et al.: Twin-robot dialogue system with robustness against speech recognition failure in human-robot dialogue with elderly people. *Appl. Sci.* **10**, 1–22 (2020). <https://doi.org/10.3390/app10041522>
14. Ma, G., Hao, Z., Wu, X., Wang, X.: An optimal electrical impedance tomography drive pattern for human-computer interaction applications. *IEEE Trans. Biomed. Circuits Syst.* **14**, 402–411 (2020). <https://doi.org/10.1109/TBCAS.2020.2967785>
15. Yadollahi, A., Shahraki, A.G., Zaiane, O.R.: Current state of text sentiment analysis from opinion to emotion mining. *ACM Comput. Surv.* **50**, 1–33 (2017)
16. Binali, H., Wu, C., Potdar, V.: Computational approaches for emotion detection in text. In: 4th IEEE International Conference on Digital Ecosystems and Technologies, pp. 172–177 (2010)
17. Canales, L., Patricio, M.-B.: Emotion detection from text: a survey. In: Proceedings of the workshop on natural language processing in the 5th information systems research working days (JISIC) (2014)
18. Kang, X., Ren, F., Wu, Y.: Exploring latent semantic information for textual emotion recognition in blog articles. *IEEE/CAA J Autom Sin* **5**, 204–216 (2018). <https://doi.org/10.1109/JAS.2017.7510421>
19. Perikos, I., Hatzilygeroudis, I.: Recognizing emotions in text using ensemble of classifiers. *Eng. Appl. Artif. Intell.* **51**, 191–201 (2016). <https://doi.org/10.1016/j.engappai.2016.01.012>
20. Nahin, A.F.M.N.H., Alam, J.M., Mahmud, H., Hasan, K.: Identifying emotion by keystroke dynamics and text pattern analysis. *Behav. Inf. Technol.* **33**, 987–996 (2014). <https://doi.org/10.1080/0144929X.2014.907343>
21. Mohammad, S.M., Kiritchenko, S.: Submitted to the Special issue on Semantic Analysis in Social Media, Computational Intelligence. Using Hashtags to Capture Fine Emotion Categories from Tweets, 1–22 (2013)
22. Uddin, M.T., Uddiny, M.A.: A guided random forest based feature selection approach for activity recognition. In: 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), pp. 1–6 (2015)
23. Sebe, N., Lew, M.S., Cohen, I., Garg, A., Huang, T.S.: Emotion Recognition Using a Cauchy Naive Bayes Classifier, 1051–465U0 (2012)
24. Vondra, M., Vich, R.: Recognition of Emotions in German Speech Using Gaussian Mixture Models. Springer LNCS (2013)
25. Bounyong, S., Yoshida, R., Yoshioka, M.: Epistemic emotion detection by video-based and heart rate variability features for online learning. In: 2023 IEEE International Conference on Consumer Electronics (ICCE), pp. 1–2 (2023)
26. Nasir, M., Dutta, P., Nandi, A.: Recognition of human emotion transition from video sequence using triangulation induced various centre pairs distance signatures. *Appl. Soft Comput.* **134** (2023). <https://doi.org/10.1016/j.asoc.2022.109971>

27. Zhou, S., Wu, X., Jiang, F., et al.: Emotion recognition from large-scale video clips with cross-attention and hybrid feature weighting neural networks. *Int. J. Environ. Res. Public Health* **20**, 1–23 (2023). <https://doi.org/10.3390/ijerph20021400>
28. Song, B.C., Kim, D.H.: Hidden emotion detection using multi-modal signals. *Ext Abstr 2021 CHI Conf Hum Factors Comput Syst* (2021)
29. Povolny, F., Matejka, P., Hradis, M., et al.: Multimodal Emotion Recognition for AVEC 2016 Challenge, 75–82 (2016)
30. Lu, Y., Zheng, W.L., Li, B., Lu, B.L.: Combining eye movements and EEG to enhance emotion recognition. *IJCAI Int. Jt. Conf. Artif. Intell.*, 1170–1176 (2015)
31. Çımtay, Y., Ekmekcioglu, E., Çağlar Özhan, Ş.: Cross-subject multimodal emotion recognition based on hybrid fusion. *IEEE Access* **8** (2020). <https://doi.org/10.1109/ACCESS.2020.3023871>
32. Poria, S., Chaturvedi, I., Cambria, E., Hussain, A.: Convolutional MKL based multimodal emotion recognition and sentiment analysis. *Proc - IEEE Int. Conf. Data Mining, ICDM* 439–448 (2017). <https://doi.org/10.1109/ICDM.2016.178>
33. Bota, P., Wang, C., Fred, A., da Silva, H.: A Review, Current Challenges, and Future Possibilities on Emotion Recognition Using Machine Learning and Physiological Signals. *IEEE Access*, 1 (2019).
34. Upreti, K., et al.: A multi-model unified disease diagnosis framework for cyber healthcare using IoMT- cloud computing networks. *J. Discrete Math. Sci. Crypt.* **26**(6), 1819–1834 (2023). <https://doi.org/10.47974/JDMSC-1831>
35. Bagga, T., Upreti, K., Kumar, N., Ansari, A.H., Nadeem, D., (eds.): *Designing Intelligent Healthcare Systems, Products, and Services Using Disruptive Technologies and Health Informatics*. CRC Press (2022)



# Explainable AI Assisted Decision-Making and Human Behaviour

Muhammad Suffian<sup>(✉)</sup> 

Department of Pure and Applied Sciences, University of Urbino, Urbino, Italy  
m.suffian@campus.uniurb.it

**Abstract.** Explainable artificial intelligence (XAI) helps users understand the logic behind machine learning model (ML) predictions so that they can better understand and believe model predictions. Many studies have looked at the interaction between humans and XAI, focusing mainly on metrics such as interpretability, fidelity, transparency, trust and usability of explanations. This paper aims to conduct a user study to explore how different types of explanations in the field of XAI affect people's understanding and behaviour in decision-making. In behavioural science, nudges and boosts are competing approaches and allow a choice architecture to improve decision-making. In our study, we utilized two types of explanations in XAI as a choice architecture, and unveiled the impactful effects of these explanations on behaviour, resulting in alternative decision-making outcomes. Explanations containing actionable information were found to be more effective and understandable. However, our findings indicate that the information provided by certain XAI techniques may not sufficiently persuade users to understand and trust the explanations offered.

**Keywords:** Explainable AI · Human Behaviour · Interpretable ML · Feature Attribution · Counterfactual Explanations

## 1 Introduction

Artificial intelligence (AI)-based algorithms (models) such as deep neural networks can make accurate predictions or decisions but often lack transparency in explaining how they arrive at those outcomes. Essentially, the internal workings of such models are not easily interpretable or explainable to humans. This problem is called *black-box* in AI. Research in the field of Explainable Artificial Intelligence (XAI) focuses on developing techniques that explain the behaviour of these black-box models in terms that humans can understand to increase the interpretability and understandability of black-box models. Miller [12] has highlighted the challenges of determining the most compelling explanation to present to a user in a given context or situation. Therefore, it is crucial to identify which types of explainable AI can improve human decision-making capabilities.

Recent research on human-AI interaction addresses how explainability affects people's perceptions, attitudes and system use, such as trust [4], behaviour and task performance [16]. However, previous research did not consider the potential

impact of explainability on users' situational information processing. For example, using currently accessible information to the user in the given context (e.g., cognitive load, facts, and knowledge) [20]. On the one hand, *feature attribution-based explanations* in XAI enables users to recognise previously unrecognised correlations between features and ground truth labels that the AI system has autonomously learned from complex data structures. This is done by showing the contribution of individual features to certain predictions [10,14]. On the other hand, *counterfactual explanations* (CE) can be used to suggest potential interventions based on the predictions of a given model. CE explains by highlighting the features (which ones and to what extent) that need to be changed to achieve the desired results [19]. For example, a loan applicant has been *Rejected*, and she is curious how the machine learning (ML) model arrived at this result. To explain this prediction, a feature attribution technique will list the most important or contributing features in the decision while a CE outlines the features with their values needed to be changed to achieve the desired outcome (*Approved*). In the recent literature on XAI, these two types of explanations are widely used for ML decisions.

We investigate which type of explanation is better understandable by users for decision-making tasks, and articulate the specific research questions to evaluate our investigation. Following are the research questions (RQs):

- **RQ 1:** How does the provision of different types of explanations impact human understanding in decision-making?
- **RQ 2:** What are the responses exhibited by individuals when presented with different types of explanations? and which was more effective and trustworthy in changing behaviour?

For the investigation and evaluation of the above RQs, the notion of soft behavioural interventions such as *nudges* and *boosts* are promising options [15] for administering the user study. Nudges do not limit options or change the underlying reward mechanism; instead, they change how choices are presented to help people make better decisions. In comparison, boosts teach people decision-making techniques that help them focus on the most important components of the option so that they can make more informed decisions (e.g. financial decisions).

We conduct a user study to empirically investigate human behaviour towards explanation types through experiments with human participants. Using the notion of nudges and boosts, we use two widely accepted explanation types, feature attribution and counterfactual explanations as *choice architecture* to explain the ML predictions and record the user responses to different asked questions (formulated based on RQs). The user responses are based on understanding 1) feature attribution-based explanations and 2) CE-based explanations. The results of our user study have shown that explanation-based AI impacts behaviour and that counterfactual explanations provide better results than a feature attribution-based technique. This study provides a foundation for future research on the influence of XAI on decision-making and possible areas of improvement in the development and deployment of explanation-enabled AI systems.

The rest of the paper is structured as follows, Sect. 2 describes the background and related work, and Sect. 3 presents the methodology for conducting the user study and experimental settings. Section 4 presents and describes the results, and Sect. 5 presents the discussion and conclusion.

## 2 Background and Related Work

Addressing the challenges of AI discussed in Sect. 1 entails a two-fold approach: firstly, understanding and assessing the user’s needs, comprehension level, and familiarity with the domain to determine the appropriate requirements and compliance of explanations; and secondly, crafting domain-specific and human-centred explanations that not only cater to the user’s understanding but also foster trust in the AI system. By prioritizing user-centric design and tailoring explanations to align with their knowledge and trust, we can pave the way for effective solutions that enhance user acceptance and confidence in AI technologies. In this regard, one of the explanation approach is **Feature Attribution-based Explanations**. Feature attribution (FA) technique is widely used and offer local and global explanations for the ML predictions, aligned with the General Data Protection Regulation (GDPR) [3] requirements. While XAI research aims to enhance explanatory power, its impact on human behaviour remains uncertain. FA approaches, like LIME (Local Interpretable Model Agnostic Explanations) [13] and SHAP (Shapley additive feature attributions) [10], quantify the significance of input variables using numerical scores. By providing FA-based explanations, AI systems can reveal undiscovered associations between features and labels extracted from complex data structures.

The other widely used explanation approaches is **Counterfactual Explanations**. Counterfactual explanations (CE) help identify interventions based on a model’s predictions [23]. CEs explain the significance of each component and suggest risk-reduction measures. For example, a machine learning model in healthcare can identify risk variables associated with a specific health condition. CEs are essential for understanding and debugging machine learning models by pinpointing inputs that lead to unexpected results and proposing improvements. They help determine which inputs caused incorrect predictions when a model produces unexpected outcomes. Generating CEs involves optimizing data points in an n-dimensional Euclidean space [23], with strategies adapting the neighbourhood space for more realistic and actionable CEs [7], sometimes involving human input [19].

When having different explanations in hand, the challenge is to identify which explanation could be utilized that can affect better on human behaviour for wider acceptance. The behavioural science provides tools like **Nudges and Boosts** to account human behaviours. Hertwig and Mazar [8] highlight the significance of nudges as essential frameworks for classifying and conceptualizing behaviour change strategies. A nudge refers to any element within the decision-making environment or choice architecture that intends to influence behaviour without imposing restrictions or mandates [21]. Nudges are designed to enhance the

decision-making structure of individuals. While boosts aim to enhance individuals' capacity to make informed and wise choices [5]. With the advancement of technology, the concept of “digital nudging” has emerged, referring to the utilization of user interface design elements in digital decision-making environments to steer people's behaviour. This adaptation of Sunstein's definition of nudge in the physical world emphasizes the digital nature of the occurrence within digital environments of such interventions [24].

Researchers have extensively studied human behaviour by presenting ML/AI predictions with and without explanations to analyze their understanding and acceptance towards both types. While the importance of XAI is already acknowledged, the focus now lies in analyzing specific explanation approaches and formats (e.g., feature attribution, counterfactual explanations) that influence human behaviour regarding trust, understanding, and acceptance. Recent research indicates that explainability generally increases perceptions of system fairness, task efficiency, human-AI cooperation, and trust in the system [25]. A systematic review [1] found that different explanations lead to varied task performance. In a study [6], ML decisions accompanied by FA score-based explanations did not improve decision-making accuracy. These findings highlight the need for further research on the effects of XAI and its various forms on human behaviour and cognitive abilities. According to the “choice architecture” hypothesis [22], how options are presented affects the decision-making process. Therefore, explanations presenting information through FA or CE must be assessed for their effectiveness, understanding, trust, and acceptance by humans.

### 3 Methods

Considering the high demand, we broadly use the term “digital nudging” to refer to situations where digital artefacts, including explanations, influence people's behaviour while respecting their interests. In this study, we employ two explanation methods, FA and CE, as a choice architecture to clarify model choices for users. This section begins by presenting the use case for the study, followed by a detailed description of the experimental design. This includes information about the participants, the user study procedure, and general demographics.

#### 3.1 Use Case Study

We selected the Bank Loan dataset<sup>1</sup> as a relevant case study, containing information about loan applicants such as income, mortgage, education, family, online account, credit card, deposit account, and security account. The task is to classify whether the loan was granted or not. This dataset is ideal due to its familiarity with people and varying features and prediction accuracy achievable by classifiers. We trained a multi-layer perceptron (MLP) neural network on this dataset, achieving a classification accuracy of 92.43% on the validation set. For

<sup>1</sup> <https://www.kaggle.com/sriharipramod/bank-loan-classification/metadata>.

a test instance representing a rejected case, we generated explanations using a FA-based explainer (LIME) [13] and a CE based explainer feedback-based counterfactual explanations (FCE) [19]. LIME with default settings, while FCE was aimed to provide customized and human-centred explanations. To prevent bias, we referred to methods A and B, respectively, in the online survey without disclosing their actual names. Similarly, to avoid the selection bias of participants from the author’s peer co-workers and social circles, we invited participants with different backgrounds and age categories. They may share similar backgrounds, interests, and perspectives, leading to a homogeneous sample that does not reflect the diversity present in the target population.

### 3.2 Experiment Design

Providing users with multiple XAI explanations reduces the effort required to understand an ML model’s prediction logic by highlighting relevant observations [11]. Therefore, we propose using at least two explanation types as a choice architecture to enhance understanding and decision-making. To align with our study’s research questions and human interaction guidelines [2], we designed two survey questions related to understanding and actionability. These questions were posed to participants during the online survey, and their responses were recorded for both explanation methods (referred to as method A and method B in the survey). The questions are as follows:

- Understanding: How well does the explanation method explain what changes are needed to obtain the model outcome as “Loan Accepted”?
- Actionability: Can a participant act upon the suggestions provided by the explanation method to change the model outcome to “Loan Accepted”?

These questions were designed to capture participants’ ratings of each explanation method’s comprehensibility and its potential for informing decision-making. The participants used a 4-point Likert scale to rate the explanations provided by each method in relation to decision-making [9]. The reason for using a 4-point Likert scale was to force the user to provide an opinion without being neutral.

Our survey design takes inspiration from user studies conducted in the field [17,18] on human evaluation of explanations. Current study’s online survey<sup>2</sup> is available at this [link](https://qfreeaccountssjcl1.a2l.qualtrics.com/jfe/form/SV_cOpboKacOifGpgy). The survey began by explaining its purpose and compliance with GDPR. Participants were asked about their familiarity with machine learning models. We introduced two types of explanations, Method-A (FA-based, i.e., LIME) and Method-B (CE-based, i.e., FCE), and described how data could be changed based on the explanations. Next, we presented a use case and explained the ML process using visual illustrations. Participants were then placed as bank loan applicants to interpret the loan rejection explanations.

Following the stage setup, participants were asked the two previously mentioned questions to record their responses regarding understanding and actionability. Subsequently, five additional questions were posed to gather participants’ opinions on both methods. The survey questions (SQ) were as follows:

<sup>2</sup> [https://qfreeaccountssjcl1.a2l.qualtrics.com/jfe/form/SV\\_cOpboKacOifGpgy](https://qfreeaccountssjcl1.a2l.qualtrics.com/jfe/form/SV_cOpboKacOifGpgy).

- SQ-1: Which explanation method was more understandable?
- SQ-2: Which explanation method provided sufficient details?
- SQ-3: Which explanation method was useful for decision-making goals?
- SQ-4: Which explanation method was more trustworthy?
- SQ-5: Which explanation method was more effective in changing participants' behaviour?

Finally, demographic data were collected with informed privacy consent. The survey was administered through [Qualtrics](https://www.qualtrics.com/)<sup>3</sup> to ensure a reliable and secure platform. To avoid confirmation bias, we referred to the explanation methods as Method-A and Method-B without explicitly mentioning their names or implementation packages (e.g., LIME, FCE). Learning effects were minimized by randomizing the order of questions and providing simple descriptions for the explanation methods.

**Participants.** The survey participants comprised the author's peer co-workers and individuals from their social circles. An anonymous link was distributed to these potential participants via email. The participants were provided with information about the anonymity of the data and were asked to participate with informed consent voluntarily. It was emphasized that the survey choices would have no real financial consequences. Participants could withdraw from the study at any time, but completing all questions was mandatory.

## 4 Results and Analysis

We obtained feedback from  $N = 42$  participants. The complete details of demographic data are provided in Table 1. No other personal data were collected at any moment during the survey. The user cannot be identified using the data from this demographic information. It was solely used to describe the sample of survey participants. The results for questions regarding understanding and actionability are plotted in Fig. 1 with bar charts. The number of participants is plotted on y-axes, and method A and method B are plotted on x-axes. The Likert scale titles are shown on the legends of the figure. Similarly, the results for five SQs are plotted in Fig. 2.

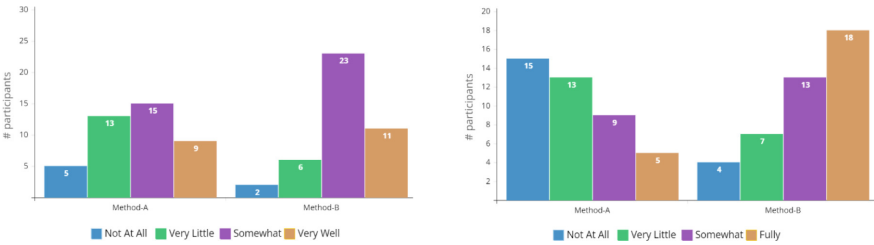
We can observe from Fig. 1 that method-B has obtained more positive responses than method-A on questions regarding understanding and actionability. Similarly, in Fig. 2, method-B have gained better results across SQs. Regarding understanding explanations (Fig. 1, left), 9 participants considered that method-A explained very well; on the contrary, 11 participants showed their consent for a very well understanding of method B. In the Likert scale (somewhat), method B has a clear majority on method-A. The second question was about the actionability (Fig. 1, right) to make a decision (to accept the explanation for its actionability). This question was designed to analyse the endorsement of the participants understanding and observe the stimuli for behaviour change.

<sup>3</sup> <https://www.qualtrics.com/>.



**Table 1.** Demographic data. The number of participants who participated and the percentage for each category are included in brackets.

Category	Sub-Category	Amount (%)
Gender	Male	27 (64%)
	Female	15 (36%)
Age	18–25	11 (26%)
	26–35	18 (43%)
	36–45	9 (21%)
	46–55	4 (10%)
Education	Upper Secondary	3 (7%)
	Bachelors (BA., B.Sc.)	7 (17%)
	Masters (M.A., M.Sc.)	13 (31%)
	Doctorate (Ph.D)	19 (45%)
Region	Europe (inc. EU, UK)	26 (62%)
	South Asia (India, Pakistan)	11 (26%)
	Japan	5 (12%)
ML Experience	No familiarity	10 (24%)
	Working level familiarity	14 (33%)
	ML experts	18 (43%)

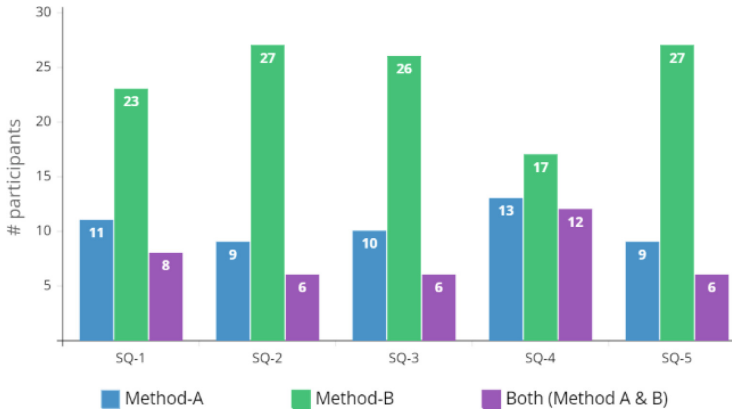


**Fig. 1.** Results of (left) understanding and (right) actionability for Method-A and Method-B.

We can observe that Method-A is losing its good score from *Not At All* to *Fully* actionable, while method-B has shown a positive growth from *Not At All* to *Fully*. These results have drawn a picture for a better understanding and actionability of Method B instead of Method-A. These results endorse our argument of choice architecture to observe the behaviour towards the explanation types.

## 5 Ethical Considerations and Limitations

According to GDPR regulations, ethical requirements for nudging interventions include either public notification or clear indication to individuals when such



**Fig. 2.** Results illustrate responses to the participants general opinion-related survey questions (SQs) for Method-A and Method-B.

interventions are used. In this study, we emphasized ethical considerations in AI by obtaining informed consent from participants, maintaining privacy protocols, and refraining from collecting personal or sensitive information. Additionally, we ensured the responsible use of nudging choices by avoiding explicit manipulation of choices to influence participant behavior. The inherent limitations of the chosen explanation methods may limit the generalizability of the study. Due to their limitations, the selected methods for feature attribution and counterfactual explanation techniques may only partially represent their respective categories. It is important to recognize these limitations and explore alternative methods to understand the findings comprehensively. Considering these factors in future works would greatly enhance the study's insights and validity.

## 6 Conclusion

In this study, we focus on presenting the choices (explanations from different methods) shown in real scenarios to observe that the user understands and can act upon those explanations. We presented two types of explanations as a choice architecture, i.e., feature attribution and counterfactual explanations. In light of our RQs, we have observed that counterfactual explanations are more stimulating in understanding and actionability, leading to behaviour change. Even though we conducted a small-scale study, it was more feasible and practical than larger studies, providing insights into human behaviour. Despite its limitations, it can provide a starting point for further research and help determine the viability of pursuing larger-scale studies. Various factors can influence the effectiveness of explanations, or their absence, in human decision-making. These factors include the specific explanation technique used, the characteristics of the dataset, and the nature of the task at hand (such as probabilistic predictions versus binary choices). Nonetheless, this paper highlights the importance of not assuming that

explainable AI will automatically enhance human decision-making. It emphasizes the necessity of assessing objective measures of human behaviour in the context of explainable AI.

## References

1. Anjomshoae, S., Najjar, A., Calvaresi, D., Främling, K.: Explainable agents and robots: Results from a systematic literature review. In: 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, 13–17 May 2019, pp. 1078–1088. International Foundation for Autonomous Agents and Multiagent Systems (2019)
2. Chromik, M., Butz, A.: Human-XAI interaction: a review and design principles for explanation user interfaces. In: Ardito, C., et al. (eds.) INTERACT 2021. LNCS, vol. 12933, pp. 619–640. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-85616-8\\_36](https://doi.org/10.1007/978-3-030-85616-8_36)
3. Donadello, I., Dragoni, M., Eccher, C.: Explaining reasoning algorithms with persuasiveness: a case study for a behavioural change system. In: Proceedings of the 35th Annual ACM Symposium on Applied Computing, pp. 646–653 (2020)
4. Erlei, A., Nekdem, F., Meub, L., Anand, A., Gadiraju, U.: Impact of algorithmic decision making on human behavior: evidence from ultimatum bargaining. In: Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, vol. 8, pp. 43–52 (2020)
5. Franklin, M., Folke, T., Ruggeri, K.: Optimising nudges and boosts for financial decisions under uncertainty. *Palgrave Commun.* **5**(1) (2019)
6. Green, B., Chen, Y.: The principles and limits of algorithm-in-the-loop decision making. *Proc. ACM Hum. Comput. Interact.* **3**(CSCW), 1–24 (2019)
7. Guidotti, R.: Counterfactual explanations and how to find them: literature review and benchmarking. *Data Min. Knowl. Disc.* 1–55 (2022). <https://doi.org/10.1007/s10618-022-00831-6>
8. Hertwig, R., Mazar, N.: Toward a taxonomy and review of honesty interventions. *Curr. Opin. Psychol.* 101410 (2022)
9. Leung, S.O.: A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. *J. Soc. Serv. Res.* **37**(4), 412–421 (2011)
10. Lundberg, S., Lee, S.I.: A unified approach to interpreting model predictions. arXiv preprint [arXiv:1705.07874](https://arxiv.org/abs/1705.07874) (2017)
11. Meske, C., Bunde, E.: Design principles for user interfaces in AI-based decision support systems: the case of explainable hate speech detection. *Inf. Syst. Front.* **25**, 743–773 (2022)
12. Miller, T.: Explanation in artificial intelligence: insights from the social sciences. *Artif. Intell.* **267**, 1–38 (2019)
13. Ribeiro, M.T., Singh, S., Guestrin, C.: Why should i trust you? Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144 (2016)
14. Ribeiro, M.T., Singh, S., Guestrin, C.: Model-agnostic interpretability of machine learning. arXiv preprint [arXiv:1606.05386](https://arxiv.org/abs/1606.05386) (2016)
15. Ruggeri, K.: Behavioral Insights for Public Policy: Concepts and Cases. Routledge, Milton Park (2018)
16. Senoner, J., Netland, T., Feuerriegel, S.: Using explainable artificial intelligence to improve process quality: evidence from semiconductor manufacturing. *Manage. Sci.* **68**(8), 5704–5723 (2022)

17. Spreitzer, N., Haned, H., van der Linden, I.: Evaluating the practicality of counterfactual explanations. In: Workshop on Trustworthy and Socially Responsible Machine Learning, NeurIPS 2022 (2022)
18. Stepin, I., Alonso-Moral, J.M., Catala, A., Pereira-Fariña, M.: An empirical study on how humans appreciate automated counterfactual explanations which embrace imprecise information. *Inf. Sci.* **618**, 379–399 (2022)
19. Suffian, M., Graziani, P., Alonso, J.M., Bogliolo, A.: FCE: feedback based counterfactual explanations for explainable AI. *IEEE Access* **10**, 72363–72372 (2022)
20. Suffian, M., Khan, M.Y., Bogliolo, A.: Towards human cognition level-based experiment design for counterfactual explanations. In: 2022 Mohammad Ali Jinnah University International Conference on Computing (MAJICC), pp. 1–5 (2022). <https://doi.org/10.1109/MAJICC56935.2022.9994203>
21. Sunstein, C.R.: Nudging and choice architecture: Ethical considerations. *Yale Journal on Regulation*, Forthcoming (2015)
22. Thaler, R.H., Sunstein, C.R.: *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Penguin, New York (2009)
23. Wachter, S., Mittelstadt, B., Russell, C.: Counterfactual explanations without opening the black box: automated decisions and the GDPR. *Harv. JL Tech.* **31**, 841 (2017)
24. Weinmann, M., Schneider, C., Brocke, J.v.: Digital nudging. *Bus. Inf. Syst. Eng.* **58**, 433–436 (2016)
25. Yang, F., Huang, Z., Scholtz, J., Arendt, D.L.: How do visual explanations foster end users' appropriate trust in machine learning? In: Proceedings of the 25th International Conference on Intelligent User Interfaces, pp. 189–201 (2020)



# Performance Measurement of Classification Algorithms for Aerial Image Registration

Hayder Mosa Merza<sup>1</sup>(✉), Ihab Sbeity<sup>1</sup>, Mohamed Dbouk<sup>1</sup>, Zein Al Abidin Ibrahim<sup>1</sup>, and Ali Salam Kadhim<sup>2</sup>

<sup>1</sup> Faculty of Sciences, Lebanese University, Beirut, Lebanon  
{hayder.merza, ihab.sbeity, mdbouk, zein.ibrahim}@ul.edu.lb

<sup>2</sup> Department of Computer Science, Baghdad University, Baghdad, Iraq  
mp.x@bk.ru

**Abstract.** The random forest algorithm is a popular machine learning technique that is widely used for classification and regression tasks. Although it is known for its high accuracy and robustness, one of the main challenges associated with the random forest algorithm is its long execution time, particularly when dealing with large datasets. Therefore, several methods have been existed to reduce the execution time of the random forest algorithm, including optimization of hyper-parameters, feature selection method and parallelization techniques. In order to do that, the Random Forest (RF) parameters have been proposed, and the experiments for tuning four RF parameters are illustrated based on CPU/GPU with PC and Nvidia Jetson board. In this paper, the results of execution time with Windows and Linux operating systems are presented. NVIDIA's Jetson platform offers great potential for embedded machine learning, aiming to strike a harmonious balance between the objectives of high accuracy, throughput and high performance. We discussed the weakness and strength for each params, and provided insights into their implementation and performance. The results of Jetson Nano board shown that the proposed methods can significantly reducing the execution time of RF params. All the coding steps are available at [https://github.com/HayderMosaMerza/image\\_registration](https://github.com/HayderMosaMerza/image_registration).

**Keywords:** Machine learning · Time series · Classification Algorithm · Parameters enhancement · CPU/GPU Nvidia Jetson Nano

## 1 Introduction

There are two major categories of artificial intelligence algorithms used for image classification: supervised and unsupervised. Supervised classifiers are most effective when there is sample training data available see Fig. 1. Another way to categorize classification algorithms is by their assumptions about the data distributed, resulting in parametric and nonparametric methods [1], since, numerous (AI) algorithms have found application across a wide range of industries and disciplines [2–4]. Model execution time is a crucial aspect of deploying machine learning models in real time word applications.

Figure 2 illustrated the cross section of Baghdad city which captured in 2007 and 2021. In this study, we discovered that reinforcing our feature space can lead to a notable enhancement in learnability. Besides this step, the performance of our predictive model testing is coming with an unsatisfactory situation, where the execute total time comes with 13 min and 50 s under one RF parameter ( $n\_jobs = 4$ ) with respect with number of trees ( $n\_estimators = 100$ ) as shown in Table 1. So, the main objective is minimizing our model execution time. The first purpose that involves construction a classification or regression rule by using (ML) algorithms with high accuracy to serve as a predictive tool for future data, like in [5] which tested five ML to prediction and evaluation, whereas the other purpose is to evaluate the relevance of potential predictor variables for the specific prediction problem at hand [6].

## 2 Related Works

The authors of [7] claimed three metrics to evaluate the image registration performance namely, inaccuracy registration, robustness method and execution time. In-Accuracy is measurement of Euclidian distance summation in every pixel between reference and sensed images positions. The statistical of inaccuracy before registration process are: mean = 4.25%, median = 2.79% and standard deviation = 5.98%. A percentage of every case which registration process improved an initial measurement of related inaccuracy is the robustness. The execution time is the computer measurement with a single CPU-thread. Authors are measured just the execution time; they didn't use load and save images time. Every method has some of parameters which impact with the performance of registration process. Setting optimal parameters is done by experts through using few of provided images.

In [8] the authors provided some of contributions for extraction-based features to detect fake news through an adaptive of new ensemble classifier model. They concerned to reducing time of classifier training. The authors are observed that prediction time of their adaptive classifier can outperforms others methods, the training time is 2.5 s. The performance is done by extracting most important features in the number of words, characters numbers, number of the sentences, average of the word length, length sentences average and recognition of name entity over ISOT and Liar datasets. Authors had clear that most of the numerous studies have limitation with a poor accuracy, the reasons can be recorded to the poor of extraction features, inefficiently parameters tuning, imbalanced of datasets etc.

The authors in [9] implemented six binary classification algorithms on two datasets, recording their performance with and without hyperparameter tuning. The algorithms used were Logistic Regression, Naïve Bayes, K-Nearest Neighbors, Decision Tree, and Random Forest. The datasets were the UCI Machine Learning Breast Cancer Wisconsin dataset and the Titanic dataset. Random state was kept constant during dataset splitting. Performance measures such as accuracy and f-score were taken from a confusion matrix. The Random Forest algorithm performed best and had no overfitting issues, with the optimal number of trees being 100 for a generalization error of less than 10%. However, increasing the number of trees may affect operational time and computational requirements.

The authors in [10] are reviewed some of various diseases which effects with crop production, and it can impact on the quantity and quality of the study area. Additionally, the identify and interpretation of wheat diseases manually will be consuming more time, also the expertise level in this field are considered as decisions related to the plants. They proposed a new model VGG19 which implement on Jetson nano GPU and obtain 99.38% of accuracy results. Also, both of CPU (Nvidia Tesla) and GPU (Nvidia K80) are used. Time performance with GPU is better than CPU which take a long time without any results.

The [11] concerned by authors to depth reconstruction as a part of computer vision. They used a single image to embedded in modern (vSLAM) vision based Simultaneous Localization and Mapping. They used PC and Nvidia Jetson with TX2 version to experiment the architecture of several (FCNNs). The suggested proposed is able running in a real time as a best performance and accuracy.

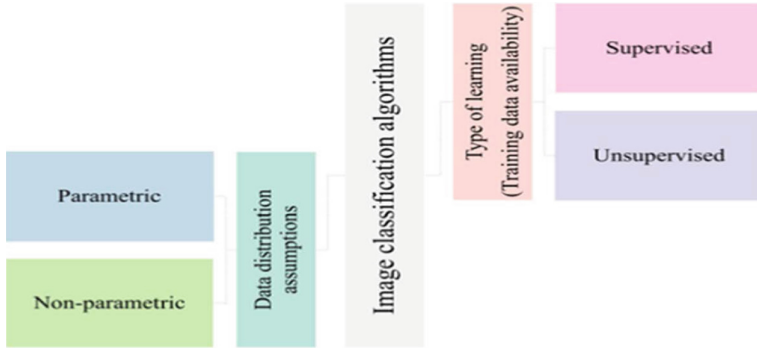
In [12], a developed version of You Only Look Once (YOLOv3) is proposed as a deep learning of object detector, it used to detect drone after transfer the learning and train the (YOLOv3). Accuracy result comes with 88.9% based on input image size with 416 x 416. The authors also concerned to detect a small, medium and large drone in a real time of small by used Nvidia Jetson (TX2).

The [13] integrates object segmentation with land cover classification method (ANN and SVM) using high-resolution images. ANN achieves higher (82.60%) accuracy than SVM (73.66%). Applying majority analysis as post-processing technique further improves average accuracy to 86.18%.

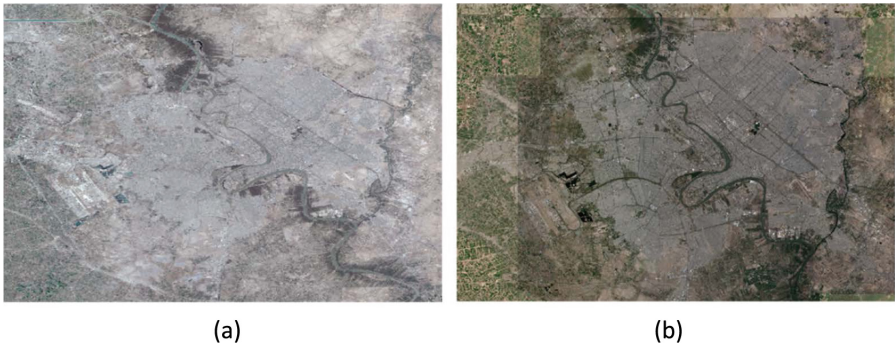
Authors claimed in [14] using a single board in Raspberry Pi4, Nvidia Jetson TX2 and Nvidia Jetson Nano via algorithms of CNN to conduct application of deep learning. They used 45K of fashion images, and the parameters that considered are (GPU, CPU, RAM and power) which covered the accuracy and the cost. The evaluated results shown that Jetson nano (TX2) is overcomes the others.

The proposed in [15] illustrated the Nvidia's Jetson platforms for machine learning to achieve the balance among the objectives of the applications for neural network (low power, high accuracy and productivity). Authors provided a survey works which evaluate the applications of neural network-based NVIDIA Jetson after discussed the optimization level of CNNs and hardware level.

The proposed in [16] demonstrated three algorithms (mono, stereo and stereo + IMU), and the performance with various visual odometry (VO) and visual inertial odometry (VIO) which evaluated on the platforms (Nvidia Jetson TX2, Xavier NX and AGX Xavier) based on the benchmark of dataset (KAIST VIO) which used for unmanned aerial vehicle (UAVs). The experiment results determined which is better algorithms and the suitable platform for UAVs system. The terms of accuracy, central processing unit (CPU) usage and the memory that used in Jetson board.



**Fig. 1.** Taxonomy of image classification algorithms.



**Fig. 2.** Baghdad city (a) 2007 (b) 2021

**Table 1.** Total time of predictive model

Algorithm	Dataset records	Operating system	Parameter1	Total Time
			n_jobs	
RF	1120000	Windows	4	13 min 50 s

### 3 Proposed Methodology

The most important note that reducing the execution time of Random Forest model may come at the cost of reduced accuracy. So, we need to find the right balance between the execution time and the accuracy of the model. We concerned by developing an integrated the multi-step methodology to prediction level, so we offered a dataset for this purpose. In order to execute the RF model, it is necessary to specify two parameters: the number of trees and the number of features randomly selected. It has been observed that RFs are more affected by variations in number of features than in number of trees [1]. Reducing the number of features parameter can accelerate the computation process, but it also diminishes the correlation between individual trees and single tree in forest.



Consequently, adjusting this parameter can have multifaceted impact on the classification accuracy [17]. Because the RF classifier is both computationally efficient and resistant to overfitting, there is no limit to the number of trees [1]. Numerous studies have identified 500 as the optimal number of trees, as increasing the number beyond this value does not lead to any significant improvement in accuracy [18]. In this study, we mainly investigated and examined four parameters, two parameters are (`n_estimators` and `max_features`) to optimize and increase predictive power of model, and (`n_jobs` and `oob_score`) are to make the model being easier to train. In addition to above, tuning time performance impacted with other parameters of RF that need to select and adjustment, each of these parameters have effect on either prediction that increase power of model, or effect on training that make the model easier to train [19].

## 4 Comparative Classifications Algorithms

To ensure the models reliability and effectiveness, we evaluate it against established classification algorithms. Figure 3 illustrates this compression based on using classification report. Here, we provide a brief overview of these algorithms:

### I. Decision Tee

DT is a versatile machine learning algorithm used for classification and regression. It splits data based on parameters, categorizing variables as “Yes” or “No” in classification and continuous in regression. DT offers advantages like suitability for both problem types, ease of use, handling of quantitative and categorical values and improved performance.

### II. Random Forest

RF is an ensemble algorithm that utilizes Decision Tree (DT) as it base, by combining multiple DT classifiers, RF effectively handles classification and regression. Each DT in RF uses its own algorithm and relies on leaf nodes. RF introduced by Ho (1995), aims for high accuracy while preventing overfitting by incorporating oblique hyper-planes, as key DT characteristics [20].

### III. Gradient Boosting

GB is a powerful machine learning algorithm used for classification and regression tasks. It is widely applied as statistical prediction model. The algorithm consists of three key steps: defining a loss function, selecting a weak learner, and creating an additive model [21].

### IV. Naïve Bayes

NB algorithm is based on probability conditions, making it straightforward. It uses a probability table as the model, updated with training data. However, NB has limitations, it is simplistic, struggles with continuous variables like time, and suitable for true online data without retraining. It involves substantial computation, especially with complex models containing numerous variables [22].

### V. Logistic Regression

LR is widely used for binomial classification tasks, providing probabilities for 0 and 1 based on inputs variables. It is applied in various scenarios like email spam classification and tumor benignity prediction. However, LR has limitations: it is not suitable for nonlinear problems, prone to overfitting, and requires identifying all relevant independent variables [22].

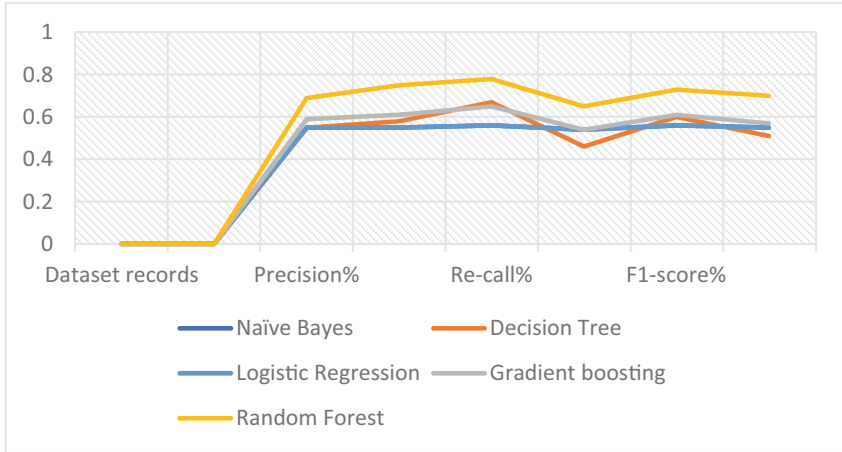


Fig. 3. Compression of Classification algorithms

### 5 Random Forest Algorithm

The Fig. 4 illustrated the general framework of machine learning-based predictive model, which involving classification analysis by analyzing historical or training data, it enables the identification of pattern that map new input records to specific output classes based on relevant independent variable values [23]. A random forest (RF) which is known as one of an ensemble classification approach that developed by Breiman [24]. As demonstrated in Fig. 5, this approach utilizes “parallel ensemble” where multiple decision tree classifiers are trained simultaneously. By doing so, it reduces the risk of overfitting and enhances prediction accuracy and control [25]. RF is an ensemble learning method in supervised learning that involves a group of tree classifiers [26]. The integration of multiple classifiers in ensemble learning can produce more dependable and trustworthy results.

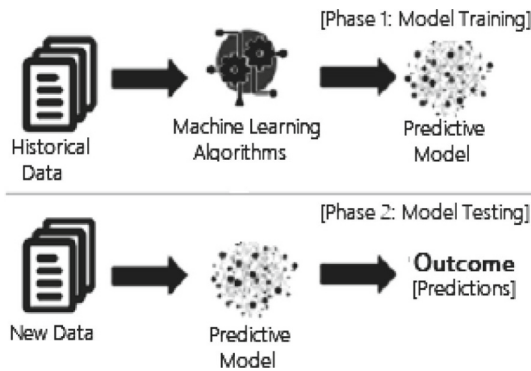


Fig. 4. Machine learning structure based predictive model and testing phase

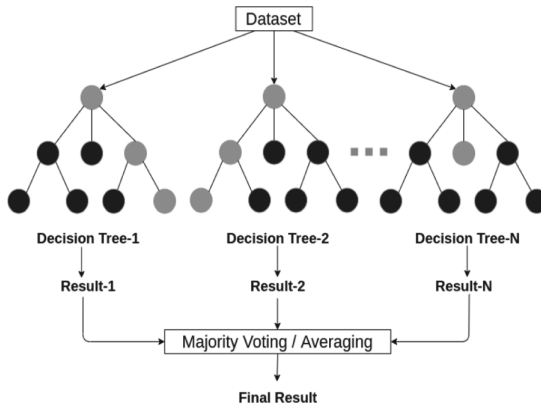


Fig. 5. An example of a random forest structure

### 5.1 RF Parameters Tuning

In RF, the Table 2 shown the numerous parameters exist that can impact classification outcomes. Each classifier in the group is built using random vector sampled independently from the input vector [27]. This technique has gained popularity in the remote sensing community as it effectively reduces of the effect of overfitting [28]. For all band combinations and tests conducted, we utilized the same parameters that demonstrated successful performance [29]. RF parameters as we can see in Fig. 6 serve two purposes: enhance the *predictive* model capability or to simplify the *training* process. Tuning process determine the optimal parameters, as well as minimizing runtime by optimization of performance measure [6]. Practically, users who use RF frequently lack confidence in whether adjusting tuning parameters to different values could enhance performance in comparison with default values that usually given by the packages of software or can computed by previous dataset [30]. RF is an algorithm that typically yields good results with default setting [31].

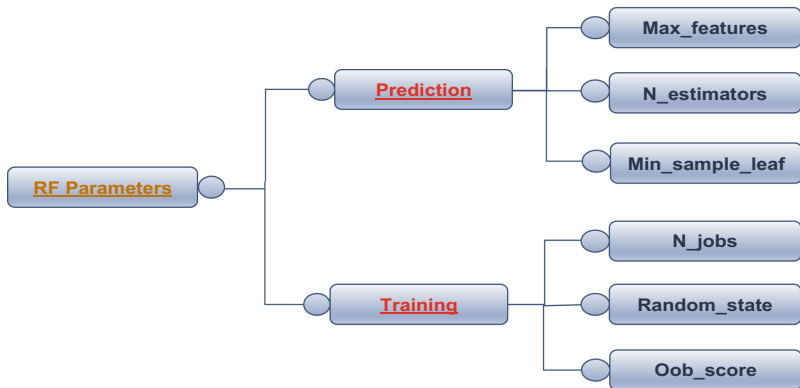
Table 2. RF Parameters

No	Random Forest Parameters	Values
1.	Bootstrap	True
2.	Class_weigh	None
3.	Critrian	Gini
4.	Max_depth	None
5.	Max_features	Log2

(continued)

**Table 2.** (continued)

No	Random Forest Parameters	Values
6.	Max_leaf_nodes	None
7.	Min_impurity_decrease	0.0
8.	Min_impurity_split	None
9.	Min_samples_leaf	1
10.	Min_samples_split	2
11.	Min_weigh_fraction_leaf	0.0
12.	N_estimators	100
13.	N_jobs	1
14.	Oob_score	False
15.	Random_state	None
16.	Verbose	0
17.	Warm_start	False

**Fig. 6.** An example of a random forest parameters

## 5.2 Prediction Model Parameters

Model's predictive power can be enhanced by primarily tuning three parameters: *N\_estimators*, *Max\_features*, and *Min\_sample\_leaf*, since number of trees (**N\_estimators**) which decide to build before taking the maximum voting or averages of predictions. A higher number of trees can lead to better performance, but it may also slow down our code. The (**max\_features**) RF can attempt a maximum number of features in an individual tree as illustrated in Fig. 7, and there are several ways to assign this limit using Python as following available options:

- Auto/None: This approach will consider all the features that are relevant in each tree, without imposing any limitations any particular tree.

- b. Sqrt: This alternative will calculate the square root of the total number of features for each run. For example, if there are 100 variables in total, only 10 of them will be considered in each tree. “log2” is another similar option for the maximum number of features.
- c. 0.2: With this choice, the RF can include 20% of the variables in each run.

In last, if it ever constructed a decision tree, it will understand the significance of the (**min\_sample\_leaf**) minimum sample of leaf size. Figure 8 present that leaf is the final node of decision tree. A smaller leaf makes the model more susceptible. Nevertheless, it should experiment with various leaf sizes to determine the optimal one for the specific situation.

### 5.3 Train Model Parameters

Several characteristics can significantly affect the speed of model training, (**n\_jobs**) parameter specifies the number of processors that engine is permitted to utilize. A value of “-1” implies that there are no limitations, while a value of “1” means that only one processor can be utilized, while (**Random state**) specific parameter can facilitate the reproducibility of a solution. By setting a fixed value, and using same parameters and training data, the same results can be obtained consistently. In (**Oob\_score**) parameter, the described technique is a cross-validation method for RF, which is considerably faster than leave-one-out validation approach. Then, for each observation, the maximum vote score is determined based solely on the tress that did not use it for training. Although a highly effective validation technique, it has drawbacks such as time-consuming and unsuitable for large datasets.

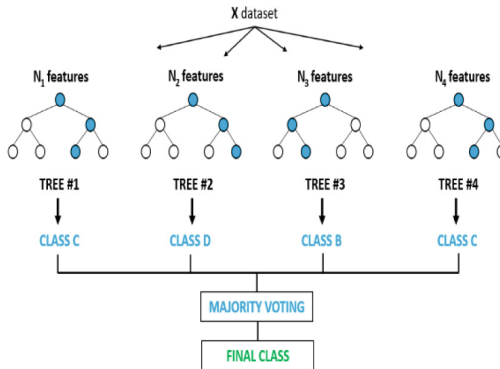
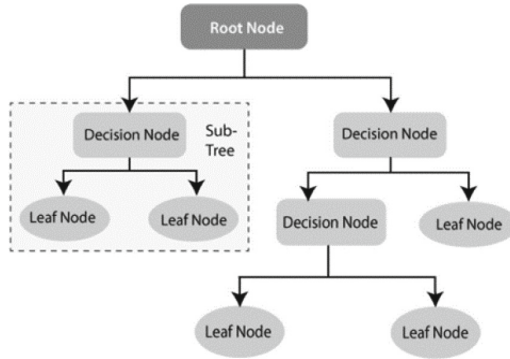


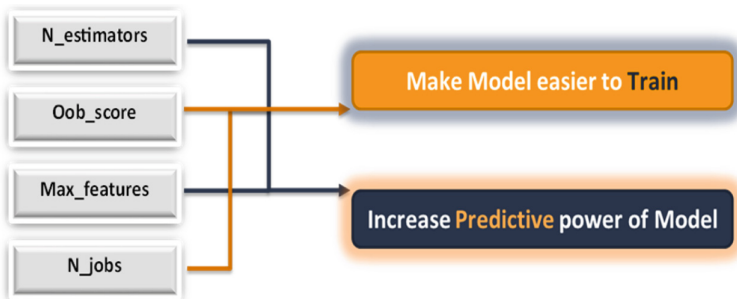
Fig. 7. Tree the number of features parameter



**Fig. 8.** An example of decision tree structure

## 6 Experiment Details

Our experiment was concerned by investigation based on tuning process of four RF parameters as shown in below Fig. 9, two predictive parameters (`n_estimators`, `max_features`) and two training parameters (`n_jobs`, `oob_score`). Table 3 shown a brief description and values of four selected parameters. The measurement excludes the time required for loading and saving images on Jupyter platform in a Windows and Linux environments by utilizing Python language version (Sect. 3.10.4). The computer hardware was configured with a 64-bit, an CPU intel iris with 8 GB RAM. In order to compute and optimize the execution time, and compare it between the performance based on two operating systems, the big and small datasets are used in step of experiment and validation, respectively. As we mentioned before, the first dataset one million and 120 thousand collected records from 294 captured images, and the second dataset is 11852 records that collected from two images for testing model time execution performance.



**Fig. 9.** Four investigated parameters of random forest

**Table 3.** Selected RF parameters

Parameters	Description	Typical default values
N_estimators	Number of trees in the forest	100, 500, 1000
Max_features	Number of features taken for better spilt at each node	None, Log2, sqrt, 0,2
Oob_score	Effective validation technique to assign used tree	True, False
N_jobs	Number of CPU cores used to paralyze tree building	-1, 1, 2, 4

## 7 Discussion

The experiments of integrated model are comes based on two phases:

### 7.1 PC-Windows and Linux Environments

Firstly, our Windows achievement is conducted on the parameters (n\_jobs, max\_features and oob\_score) with considering (n\_estimators = 100). The below Tables 4, 5 and 6 shown the time performance results based on tuning the mentioned three RF parameters. We can observe in Table 5 the time performance comes with lower execution time which is better than Tables 7, 8 and 9 that presented based Linux system. Secondly, max\_feature parameter takes another value in additional to (log2), which is (sqrt) value. Hence, it can notice in Tables 10 and 11 and Fig. 10 that applied log2 in Windows operating system is better than (sqrt) value in both Windows and Linux operations. Finally, for validation step, we utilized two images to compute the registration and execution time process. In Tables 12, 13 and 14, the results were satisfied, we can conclude that enabling (oob\_score/True-False) parameter negatively affects the model’s performance time. Therefore, we should neglect it. Additionally, based on all results, we can see that adjusting the parameters of the random forest model can improve its performance. Specially, after implementing a new dataset, as you can see Table 13, the adjusted parameters time result in a performance time is 141 ms.

**Table 4.** Total time for one parameter of RF in Windows (n\_estimators = 100)

Algorithm	Dataset records	Operating system	Parameter 1	Total Time
			n_jobs	
RF	1120000	Windows	4	13min 50s
RF	1120000	Windows	2	12min 53s
RF	1120000	Windows	1	10min 49s
RF	1120000	Windows	-1	23min 45s

**Table 5.** Total time for two parameters of RF in Windows ( $n_{estimators} = 100$ )

Algorithm	Dataset records	Operating system	Parameter1	Parameter 2	Total Time
			n_jobs	Max_features	
RF	1120000	Windows	4	Log2	13min 26s
RF	1120000	Windows	2	Log2	9min 6s
RF	1120000	Windows	1	Log2	7min 12s
RF	1120000	Windows	-1	Log2	19min 26s

**Table 6.** Total time for three parameters of RF in Windows ( $n_{estimators} = 100$ )

Algorithm	Dataset records	Operating system	Parameter1	Parameter2	Parameter3	Total Time
			n_jobs	Max_features	Oob_score	
RF	1120000	Windows	4	Log2	True	13min 53s
RF	1120000	Windows	2	Log2	True	9min 8s
RF	1120000	Windows	1	Log2	True	7min 42s
RF	1120000	Windows	-1	Log2	True	20min 2s

**Table 7.** Total time for one parameters of RF in Linux ( $n_{estimators} = 100$ )

Algorithm	Dataset records	Operating system	Parameter 1	Total Time
			n_jobs	
RF	1120000	Linux	4	21min 18s
RF	1120000	Linux	2	21min 26s
RF	1120000	Linux	1	20min 30s
RF	1120000	Linux	-1	22min 9s

**Table 8.** Total time for two parameters of RF in Linux ( $n_{estimators} = 100$ )

Algorithm	Dataset records	Operating system	Parameter 1	Parameter 2	Total Time
			n_jobs	Max_features	
RF	1120000	Linux	4	Log2	15min 23s
RF	1120000	Linux	2	Log2	15min 43s
RF	1120000	Linux	1	Log2	16min
RF	1120000	Linux	-1	Log2	15min 35s



**Table 9.** Total time for three parameters of RF in Linux (n\_estimators = 100)

Algorithm	Dataset records	Operating system	Parameter 1	Parameter 2	Parameter 3	Total Time
			n_jobs	Max_features	Oob_score	
RF	1120000	Linux	4	Log2	True	16min 5s
RF	1120000	Linux	2	Log2	True	15min 57s
RF	1120000	Linux	1	Log2	True	15min 20s
RF	1120000	Linux	-1	Log2	True	16min 50s

**Table 10.** Total time for two parameters of RF with sqrt in Windows (n\_estimators = 100)

Algorithm	Dataset records	Operating system	Parameter 1	Parameter 2	Total Time
			n_jobs	Max_features	
RF	1120000	Windows	4	sqrt	14min 26s
RF	1120000	Windows	2	sqrt	12min 56s
RF	1120000	Windows	1	sqrt	11min 31s
RF	1120000	Windows	-1	sqrt	29min 49s

**Table 11.** Total time for two parameters of RF with sqrt in Linux (n\_estimators = 100)

Algorithm	Dataset records	Operating system	Parameter 1	Parameter 2	Total Time
			n_jobs	Max_features	
RF	1120000	Linux	4	sqrt	24min 33s
RF	1120000	Linux	2	sqrt	24min 34s
RF	1120000	Linux	1	sqrt	14min 56s
RF	1120000	Linux	-1	sqrt	27min 48s

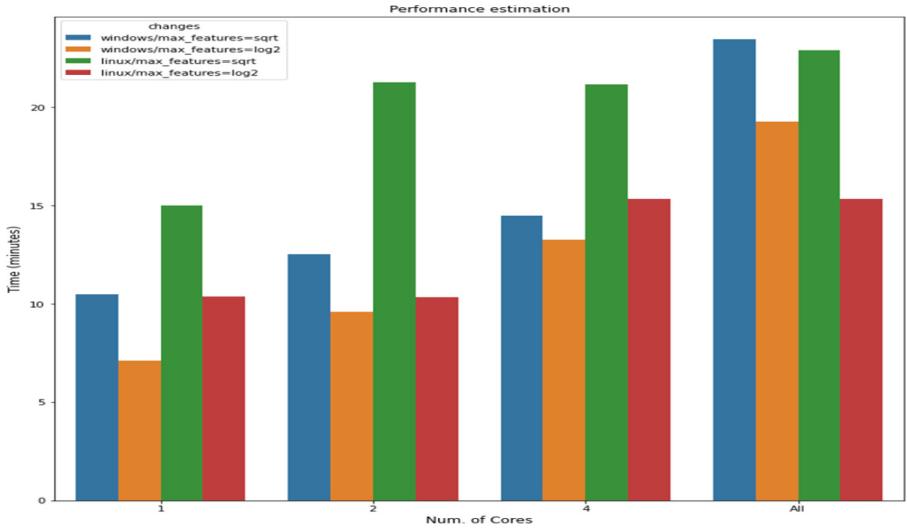


Fig. 10. Two max\_features values in Windows and Linux

Table 12. Performance time for one parameter of RF in Windows (n\_estimators = 100)

Algorithm	Dataset		Dataset records	Operating system	Parameter 1	Total Time
	Original	Training			n_jobs	
RF	Baghdad2	Baghdad2_2	11852	Windows	4	312ms
RF	Baghdad2	Baghdad2_2	11852	Windows	2	188ms
RF	Baghdad2	Baghdad2_2	11852	Windows	1	172ms
RF	Baghdad2	Baghdad2_2	11852	Windows	-1	219ms

Table 13. Performance time for two parameters of RF in Windows (n\_estimators = 100)

Algorithm	Dataset		Dataset records	Operating system	Parameter 1	Parameter 2	Total Time
	Original	Training			n_jobs	Max_features	
RF	Baghdad2	Baghdad2_2	11852	Windows	4	Log2	219ms
RF	Baghdad2	Baghdad2_2	11852	Windows	2	Log2	219ms
RF	Baghdad2	Baghdad2_2	11852	Windows	1	Log2	141ms
RF	Baghdad2	Baghdad2_2	11852	Windows	-1	Log2	234ms

**Table 14.** Performance time for three parameters of RF in Windows ( $n_{estimators} = 100$ )

Algorithm	Dataset		Dataset records	Operating system	Parameter1	Parameter 2	Parameter3	Total time
	Original	Training			n_jobs	Max_features	Oob_score	
RF	Baghdad2	Baghdad2_2	11852	Windows	4	Log2	True	484 ms
RF	Baghdad2	Baghdad2_2	11852	Windows	2	Log2	True	375 ms
RF	Baghdad2	Baghdad2_2	11852	Windows	1	Log2	True	328 ms
RF	Baghdad2	Baghdad2_2	11852	Windows	-1	Log2	True	344 ms

## 7.2 Nvidia Jetson Nano-CPU/GPU

This experiment was implemented on a jetson Nano machine computing and obtained an optimal accuracy of classification process. By parameters tuning which implemented on Linux environment of Jetson nano, the performance of our model can be enhanced. Our main purpose has been concern to assess GPU performance accelerator over 128-GPU processor. This step a Jetson Nano toolkit is developed which was used for training model. Due to the type and size of data, the 4 cores-CPU processor taken long model training time, and in some of cases, it returned without any output results. The architecture of GPU was an NVIDIA Maxwell with 128 NVIDIA CUDA cores, Quad-Core ARM Cortex-A57 Processor of CPU, and 4GB LPDDR4 Memory, the main storage is MicroSD Slot, and the Gigabit Ethernet Port was used for connectivity. The Table 15 comes with experiment based on 1 million and 120K dataset records, it can observed the difference of execution time which Jetson Nano adopted. Table 16 illustrated the performance of execution time based on the same two images (Baghdad2) the reference and sensed, it is clear that performance is better than applied on PC-Windows environment. For the last step of performance, undoubtedly, the Jetson board-GPU is overcome the PC-CPU by getting on negligible performance time results, see the Table 17.

**Table 15.** Total time for RF two parameters with CPU in Nvidia Jetson Nano ( $n_{estimators} = 100$ )

Algorithm	Dataset records	Nvidia Jetson Nano-OS	Processor	Parameter 1	Parameter 2	Total Time
				n_jobs	Max_features	
Random Forest	1120000	Linux	CPU	1	Log2	19.8 ms

**Table 16.** Performance time for RF two parameters with CPU in Nvidia Jetson Nano (n\_estimators = 100)

Algorithm	Dataset records	Nvidia Jetson Nano-OS	Processor	Parameter 1	Parameter 2	Total Time
				n_jobs	Max_features	
Random Forest	11852	Linux	CPU	1	Log2	18.8 ms

**Table 17.** Performance time for RF two parameters with GPU in Nvidia Jetson Nano (n\_estimators = 100)

Algorithm	Dataset records	Nvidia Jetson Nano-OS	Processor	Parameter 1	Parameter 2	Total Time
				n_jobs	Max_features	
Random Forest	11852	Linux	GPU	1	Log2	17.6 ms

## 8 Conclusion and Future Work

RF algorithm is the one of widespread and strong of feature selection which used in multi different fields. It can improve classification accuracy through selecting important features. Despite the effectiveness of our RF model, it includes some of shortcomings which related with accuracy, features selected number and the execution time. In this study, we applied one of the best ways to reduce execution time, also we conducted and presented the experiment details for model performance enhancement. Enhancement process is done by select four parameters RF tuning with consideration right balance between accuracy and execution time. The train and predictive parameters which we employed are (n\_estimators, n\_jobs, max\_features and oob\_score). The Windows and Linux was implemented operating system, and the adjustment process for RF parameters has been with CPU on PC and GPU on the board Nvidia Jetson nano. As the future work, it can decrease the number of trees in the forest, the model becomes simpler and faster. However, it might decrease the accuracy of the model. Also, by decreasing the number of features to train the model, the execution time can be reduced. It can use feature selection methods such as feature importance or mutual information to determine the most relevant feature.

## References

1. Sheykhmousa, M., Mahdianpari, M., Mohammadimanesh, F., Ghamisi, P., Homayouni, S.: Support vector machine versus random forest for remote sensing image classification: a meta-analysis and systematic review. *IEEE J. Sel. Top. App. Earth Obser. Remo. Sens.* **13**, 6310 (2020)

2. Mutlu, E.N., Devim, A., Hameed, A.A., Jamil, A.: Deep Learning for Liver Disease Prediction 13 (April 2022)
3. Mercan, V., et al.: Hate speech and offensive language detection from social media. In: 2021 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube) (26 October 2021)
4. Rasheed, J., et al.: A survey on artificial intelligence approaches in supporting frontline workers and decision makers for the COVID-19 pandemic. Elsevier **141**, 110337 (1 December 2020). <https://doi.org/10.1016/j.chaos.2020.110337>
5. Guney, S., Kilinc, I., Hameed, A.A., Jamil, A.: Abalone age prediction using machine learning. In: Pattern Recognition and Artificial Intelligence (13 April 2022). [https://doi.org/10.1007/978-3-031-04112-9\\_25](https://doi.org/10.1007/978-3-031-04112-9_25)
6. Probst, P., Wright, M.N., Boulesteix, A.-L.: Hyperparameters and tuning strategies for random forest. WIREs Data Mining Knowl Discov. (2019). <https://doi.org/10.1002/widm.1301>
7. Borovec, J., Munoz-Barrutia, A., Kybic, J.: Benchmarking of image registration methods for differently stained histological slides. IEEE no. 978-1-4799-7061-2/18/\$31.00 ©2018 IEEE
8. Hakak, S., et al.: An ensemble machine learning approach through effective feature extraction to classify fake news. Future Generation Computer Systems 47–58 (2021)
9. Bahel, V., Pillai, S., Malhotra, M.: A comparative study on various binary classification algorithms and their improved variant for optimal performance. IEEE, 5–7 (2020). no. 978-1-7281-7366-5/20/\$31.00 ©2020
10. Aboneh, T., Rorissa, A., Srinivasagan, R., Gemechu, A.: Computer vision framework for wheat disease identification and classification using jetson GPU infrastructure. Technologies **9**, 47 (2021). <https://doi.org/10.3390/technologies9030047>
11. Bokovoy, A., Muravyev, K., Yakovlev, K.: Real-time Vision-based Depth Reconstruction with NVidia Jetson. IEEE (2019). 978-1-7281-3605-9/19/\$31.00
12. Wei Xun, D.T., Lim, Y.L., Srigrarom, S.: Drone detection using YOLOv3 with transfer learning on NVIDIA Jetson TX2. In: IEEE 2021 Second International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP), no. <https://doi.org/10.1109/ICA-SYMP50206.2021.9358449>
13. Jamil, A., Khan, A.A., Hameed, A.A., Baza, S.U.: Application of machine learning approaches for land cover classification. J. Appl. Emer. Sci. **11**(1) (2021)
14. Süzen, A.A., Duman, B., Şen, B.: Benchmark Analysis of Jetson TX2, Jetson Nano and Raspberry PI using Deep-CNN. IEEE (2020)
15. Mittal, S.: A Survey on optimized implementation of deep learning models on the NVIDIA Jetson platform. J. Syst. Architect. **97**, 428–442 (2019). <https://doi.org/10.1016/j.sysarc.2019.01.011>
16. Jeon, J., et al.: Run Your Visual-Inertial Odometry on NVIDIA Jetson: Benchmark Tests on a Micro Aerial Vehicle. IEEE Robo. Autom. Lett. **6**(3) (2021)
17. Klusowski, M.: Complete Analysis of a Random Forest Model **13**, 1063–1095 (2018). no. [arXiv:1805.02587](https://arxiv.org/abs/1805.02587)
18. Belgiu, M., Drăgu, L.: Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens. **114**, 24–31 (2016)
19. Srivastava, T.: Analytics Vidhya, Tuning the parameters of your Random Forest model (26 June 2020). [Online]. Available: <https://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/>. Accessed 9 June 2015
20. Dai, B., Chen, R.-C., Zhu, S.-Z., Zhang, W.-W.: Using Random Forest Algorithm for Breast Cancer Diagnosis. International Symposium on Computer, Consumer and Control (IS3C) (2018). <https://doi.org/10.1109/is3c.2018.00119>
21. Natekin, A., Knoll, A.: Gradient boosting machines, a tutorial. Frontiers in Neurobotics **7** (December 2013). <https://doi.org/10.3389/fnbot.2013.00021>

22. Ray, S.: A quick review of machine learning algorithms. In: International Conference on Machine Learning, Big Data, Cloud and Parallel Computing. IEEE, Faridabad, India (2019)
23. Gondia, A., Siam, A., El-Dakhkhni, W., Nassar, A.H.: Machine learning algorithms for construction projects delay risk prediction. American Society of Civil Engineers (2020). [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001736](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001736)
24. Breiman, L.: ST4\_Method\_Random\_Forest. Mach. Learn. **45**, 5–32 (2001)
25. Sarker, I.H.: Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science (2021). <https://doi.org/10.1007/s42979-021-00592-x>
26. Pal, M., Mather, P.: Support vector machines for classification in remote sensing. Int. J. Remote Sens. **26**, 1007–1011 (2005)
27. Pal, M.: Random forest classifier for remote sensing classification. Int. J. Remote Sens. **26**, 217–222 (2005)
28. Belgiu, M., Drăguț, L.: Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens **114**, 24–31 (2016)
29. Wessel, M., Brandmeier, M., Tiede, D.: Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on Sentinel-2 Data. Remote Sens. **10**, 1419 (2018). <https://doi.org/10.3390/rs10091419>
30. Probst, P., Bischl, B., Boulesteix, A.-L.: Tunability: Importance of hyperparameters of machine learning algorithms (2018). ArXiv preprint, no. Retrieved from <https://arxiv.org/abs/1802.09596>
31. Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D.: Do we need hundreds of classifiers to solve real world classification problems. J. Mach. Learn. Res. **15**, 3133–3181 (2014)



# Investigating IoT-Enabled 6G Communications: Opportunities and Challenges

Radia Belkeziz<sup>1</sup>, Reda Chefira<sup>1,2</sup>(✉), and Oumaima Tibssirte<sup>1</sup>

<sup>1</sup> Private University of Marrakesh, Marrakesh, Morocco  
{r.belkeziz, r.chefira, o.tibssirte}@upm.ac.ma

<sup>2</sup> Cadi Ayyad University, Marrakesh, Morocco

**Abstract.** IoT has gained considerable momentum over the years. The development of standards and protocols has enabled the growth of huge and varied applications. Starting with an overview of the current landscape of IoT communication protocols, the study highlights the significance of protocols as a means of enabling seamless connectivity and data exchange between IoT devices. It then examines the opportunities and challenges of IoT-based 6G, featuring existing IoT communication and messaging protocols. Further, the paper dives into technologies and architectures that support IoT communication and messaging, while surveying Artificial Intelligence paradigms backing prospective applications. This investigation supplies a foundation for ongoing research and development designed not only to enhance IoT communication, but also to drive coverage and maximize the potential of IoT enabled applications across a range of industries.

**Keywords:** IoT communication · terrestrial network · satellite network · IST-network · machine learning · artificial intelligence · computing systems · blockchain

## 1 Introduction

The Internet of Things (IoT) has transformed the way devices and systems communicate and interact with one another, allowing for a wide range of applications and services. The communication protocols that enable smooth connectivity and data sharing between devices are at the heart of IoT. Because they specify the rules and standards for delivering information, IoT communication protocols are critical in ensuring efficient and secure communication. In this article, we will look at the importance of existing IoT communication protocols and networking communication technologies such as Bluetooth, Z-wave, 6LoWPAN, ZigBee, and many others. Further, this paper highlights messaging protocols such as HTTP, MQTT, CoAP, XMPP, and AMQP and gives the basic properties of each one.

As the IoT ecosystem grows, there is a greater demand for dependable and ubiquitous connections in a variety of locations, including distant and rural areas, maritime applications, and aviation industries. While terrestrial networks are the backbone of IoT connectivity, they frequently have coverage restrictions, particularly in distant areas

where infrastructure installation is difficult. Satellite networks come in handy here, giving wide-area coverage and worldwide communication.

Because satellite and terrestrial networks are integrated, IoT devices may smoothly switch between network technologies, maximizing connectivity based on availability, cost, and performance. This connection enables a more robust and durable Internet of Things infrastructure, ensuring continuous connectivity even in the most remote areas.

The convergence of satellite and terrestrial networks is likely to become even more important with the introduction of next-generation networks such as 6G. 6G seeks to transform wireless communication by delivering ultra-low latency, large device connectivity, and tremendous bandwidth. IoT devices can benefit from both satellite and terrestrial networks by integrating them into the 6G ecosystem. Satellites can provide global coverage and support applications such as asset tracking, environmental monitoring, and disaster management that require continuous communication. Terrestrial networks, on the other hand, can serve applications that demand ultra-low latency, such as driverless vehicles and real-time industrial control systems, and can provide high-speed connectivity in crowded urban areas.

With the merging of satellite and terrestrial networks in 6G, new opportunities and difficulties for IoT communication will emerge. It would necessitate flawless compatibility of various network technologies, as well as effective handover mechanisms and intelligent network management strategies. Furthermore, creative techniques to address security, privacy, and scalability challenges in this hybrid network environment will be required.

Furthermore, in the context of satellite and terrestrial network integration, AI can be critical in optimizing network resources, managing the transition between different network technologies, and intelligently routing data based on network conditions and device requirements. To dynamically assign resources, predict network congestion, and optimize network performance, AI algorithms can examine real-time data from IoT devices, satellites, and terrestrial networks.

Additionally, AI-powered edge computing capabilities can help reduce latency and enable real-time decision-making at the network's edge. This is especially important for time-sensitive IoT applications like autonomous vehicles and industrial automation systems that require fast reactions and low-latency connectivity. IoT devices can analyze data locally by putting AI algorithms and models at the network edge, decreasing the requirement for data transmission to the cloud or centralized servers, reducing latency, and enhancing overall system performance. In this paper, we discuss the opportunities and challenges of the IoT driven 6G communication by resending the technologies and service architectures and discussing its opportunities and challenges.

The remainder of this paper is as follows: Sect. 2 presents existing IoT communication protocols while Sect. 3 is about its messaging protocols. Section 4 introduces the integration of satellite and terrestrial networks through their evolution across generations. Section 5 presents the emerging AI-based 6G challenges for STN. Section 6 introduces service architectures for STN IoT. Section 7 discusses the presented architectures and provides an overall summary. Finally, Sect. 8 summarizes the paper through a conclusion.



## 2 Communication Protocols

This section introduces the IoT protocols, as standards and methods that enable data and information exchange between devices, which also allow the transmission of data between internet and devices at the edge in a way that systems communicate properly.

### 2.1 Bluetooth Classic/BLE

The Bluetooth Classic which is basically the main radio protocol behind wireless speakers, headphones, car entertainment systems [1], is a low power protocol that streams data over 79 channels in the 2.4GHz unlicensed industrial, scientific and medical (ISM) frequency band, supporting point-to-point device communication [2]. Bluetooth Low Energy (BLE) is a kind of Bluetooth technology which is specified for very low power operation [3]. Transmitting data over 40 channels compared to classic Bluetooth, it expands from point-to-point to broadcast supporting large-scale device networks [4].

### 2.2 ZigBee

ZigBee is a standard developed by the ZigBee Alliance for Personal-Area Networks [5]. Zigbee is a wireless communication protocol used in the connected home. Working in a mesh network, it can connect more than 65 000 devices. ZigBee includes a suite of the standard IEEE 802.15.4 specifications communication protocols enabling low-power wireless personal area networks (WPAN) for point-to-point multi-topology communication, point-to-point and multi-point-to-point between devices [6]. Extensions to this standard cover encryption, data authentication, routing and transmission. Features of Zigbee are remote control, intelligence-enhanced energy monitoring, automated building systems, smart healthcare, smart lighting, network peripherals and numerous Intelligent Network Services [7].

### 2.3 Z-wave

Specifically used in smart home networks, Z-wave is a wireless communication protocol designed for smart devices to network and share data and control commands within smart home networks [8]. This technology may be used to control lights, heating and air conditioning, and appliances and home security, among other functions [9]. Z-wave operates in a 902–928 MHz band, which does not interfere with wi-fi and other common wireless transmission. It can be used as an alternative to Zigbee. It is an active technology, simple to install and use [7].

### 2.4 Wi-Fi

Using radio waves, Wi-Fi technology transmits data from a wireless router to compatible devices such as computers, cell phones and other equipment [7]. It simply allows these devices to exchange information with one another, creating a network and such devices rely on standard wireless technology for Internet connectivity. In fact, these standards evolve regularly. Updating helps speed up Internet access and increase the number of simultaneous connections.

## 2.5 NFC

NFC, or Near Field Communication, is a technology for exchanging data between a reader and any compatible mobile terminal or between the terminals themselves [10]. Since its release, NFC has been used in more and more different applications, for security measures, convenience, and even transactions, and to automate home environments for a user by locking and unlocking doors, lighting, air conditioning, personal computers etc. [11]. NFC makes connecting two separate IoT devices easy and intuitive thanks to its simple tap-and-go mechanism. It also provides data security on many levels, such as, NFC chips must be in near proximity to each other to initiate a transaction. This defends against hackers gaining unauthorized entry.

## 2.6 LoWPAN

LoWPAN is an IP-based standard internetworking protocol and is considered a leading standard for IoT communication protocols [12]. It relies on IP communication over low power wireless IEEE 802.15.4 networks utilizing IPv6. 6LoWPAN provides a means of carrying packet data in the form of IPv6 over IEEE 802.15.4 and other networks [13]. It provides end-to-end IPv6 and as such it is able to provide direct connectivity to a huge variety of networks including direct connectivity to the internet.

## 2.7 LoRAWAN

To be able to define LoRAWAN technology, we first define LoRA, which is a wireless audio frequency spectrum. It's a physical layer protocol that uses spread spectrum modulation and supports long-range communication at the cost of a narrow bandwidth [14]. LoRAWAN is simply a low-power wide area networking protocol built on top of the LoRA radio modulation technique [15]. It wirelessly connects devices to the internet and manages communication between end-node devices and network gateways [14].

## 2.8 Cellular

Cellular IoT is the technology that connects physical objects to the internet utilizing the same cellular network currently used by smartphones. Leveraging GSM/3G/4G cellular communication capabilities, it supplies safe, real-time, broadband connectivity to the Internet [16]. In other words, it can connect IoT devices using existing mobile networks. Cellular technology is suitable when high data rates are required, and applications need to operate over long distances [17].

# 3 Messaging Protocols

IoT messaging protocols define a structured method for communication on the internet. It allows applications to exchange payloads of data. HTTP, MQTT, CoAP XMPP and AMQP are protocols to describe how an implementation should process messages and route data between producers and consumers.

### 3.1 HTTP

Hypertext Transfer Protocol (HTTP) is a protocol designed to transmit hypermedia documents, such as HTML, over the application layer [18]. HTTP is the most common protocol that is used for IoT devices when there is a lot of data to be published. However, the HTTP protocol is not preferred because of its cost, battery-life, energy saving, due to TCP execution and operation via a three-way handshake process. HTTP does not lend itself well to low-energy embedded processes. REST methods are associated with it, using the GET, POST, PUT and DELETE methods for updating, creating, reading and deleting transactions [19].

### 3.2 MQTT/SMQTT

IoT communication protocol MQTT is increasingly popular as a publish/subscribe messaging protocol [20]. The entire set of devices interacts via a broker or server. Client devices may receive or accept data directly from the broker, through subscription to specific topics. MQTT is often criticized due to its lack of security. SMQTT stands for Secure-MQTT [21]. It used lightweight attribute-based encryption. SMQTT is proposed only to enhance MQTT security features. SMQTT is a session layer protocol. It has a broadcast encryption feature which encrypts one message and delivers it to multiple nodes. Before publishing, data is being encrypted. Then after the broker publishes it and when received, subscribers decrypt data with the same master key generated at the beginning by the algorithm chosen by developers.

### 3.3 CoAP

Constrained Application Protocol (CoAP) is a specialized web transfer protocol that is designed to enable simple, constrained devices to join the IoT even through constrained networks with low bandwidth and low availability [18]. It functions as a HTTP for restricted devices, enabling equipment such as sensors or actuators to communicate on the IoT. These sensors and actuators are controlled and contribute by passing along their data as part of a system [8]. The protocol is designed for reliability in low bandwidth and high congestion through its low power consumption and low network overhead.

### 3.4 XMPP

Extensible Messaging and Presence Protocol (XMPP) provides an XML-based communication protocol designed for message-driven middleware and terminal applications including real-time instant messaging and time attendance information [22]. Built to be scalable, XMPP features publishing and subscription systems, file transfer and embedded IoT network communication. The XML text format makes interfacing M2M IoT networks with M2P communications, whenever applicable, straightforward.

### 3.5 AMQP

Advanced Message Queuing Protocol (AMQP) represents asynchronous message queuing [18]. It is a binary application-layer protocol suitable for middleware oriented towards messages. It deals with publishers and consumers. The publishers produce the messages, the consumers pick them up and process them. The message broker is responsible for ensuring that the messages from a publisher go to the right consumers [8]. To do that, the broker uses two key components: exchanges and queues. When a Publisher delivers messages to a Designate Exchange, a Consumer either pulls messages from a Queue, or, depending on the configuration, the Queue pushes them to the Consumer. After discussing the significance of IoT communication protocols, we will now focus on the integration of satellite and terrestrial networks, which is critical in expanding the reach and capabilities of IoT connectivity in the next chapter.

## 4 Satellite and Terrestrial Network Integration

Many investigations have been done on an integrated satellite-terrestrial (IST) IoT network, where satellite networks (SN) provide wireless services to the IoT devices uncovered by terrestrial cellular networks. The integration of satellite and terrestrial networks (SN) is an opportunity not only to ensure network connectivity to uncovered areas, but it could be a way to guarantee connectivity to IoT applications that handle sensible or critical data in addition to natural disasters scenarios. Many challenges still face this opportunity and there are several research that addresses some propositions. The Integration of satellite and terrestrial networks across generations is detailed as follows:

- 3rd Generation
  - For the Internet of Things (3G), there was little terrestrial and satellite network integration. Satellite networks mainly offered coverage in remote locations without adequate terrestrial infrastructure, but there was little integration between the two.
  - IoT Impact: In remote or rural areas, 3G-enabled IoT devices had few connectivity options. Applications for the Internet of Things that relied on satellite communication were frequently stand-alone and did not seamlessly integrate with terrestrial networks.
- 4th Generation
  - Integration: With 4G, more effort was put into closely integrating satellite and terrestrial networks. IoT devices can now switch between terrestrial and satellite networks based on availability thanks to hybrid network solutions.
  - IoT Impact: IoT devices had better connectivity options and could use satellite networks in areas with insufficient terrestrial coverage. IoT applications needing extended coverage in off-the-grid or difficult environments benefited from this integration.
- 5th Generation
  - Integration: For seamless connectivity, 5G promises a continuous enhancement of terrestrial integration with satellite networks. To bring 5G to rural areas and support IoT applications in various settings, satellite networks were essential.
  - Impact on IoT: The increased coverage offered by satellite integration may be advantageous for IoT devices utilizing 5G networks. This made it possible for IoT applications in sectors like agriculture, shipping, and transportation to function dependably and effectively over a wide range of geographical locations.

- 6th Generation
    - Satellite and terrestrial network integration is anticipated to be more seamless and effective in 6G. The objective is to achieve ubiquitous connectivity, enabling seamless network switching for IoT devices.
    - Impact on IoT: By combining satellite and terrestrial networks, 6G will allow IoT devices to stay connected in a variety of environments and maintain constant communication. This integration will be useful for IoT applications running in remote areas, on aerial platforms, or in places with patchy terrestrial coverage.
- In summary, the integration of satellite and terrestrial networks for IoT has significantly improved over the years, going from a basic level of integration in 3G to a more sophisticated level in later generations. The integration has improved coverage, dependability, and seamless connectivity for IoT devices with each generation. For IoT applications running in difficult environments and remote locations, 6G will put a strong emphasis on achieving ubiquitous connectivity and smooth transitions.

## 5 Emerging AI-based 6G Solutions for STN

The need for AI to improve the quality, reliability, and service efficiency of continuity and ubiquity of STN systems demands increased R&D attention. Recent research challenges of AI, ML and DL-based solutions and frameworks mainly focus on:

### 5.1 Network Optimization

Satellite networks are key components of the global network. Yet, several issues arise in the development of routing strategies such as the design of a dynamic architecture and a capable topology to ensure inter-satellite connection [23]. Indeed, users' mobility and volatile environment cause the network topology to change. Using AI can assist in developing strategies to predict real-time traffic needs according to the network environment [24, 25].

### 5.2 QoS

To benchmark the performance of communication networks, QoS is the first parameter to be used. In [26, 27], several papers are examined dealing mainly with wireless transmission, resource management and allocation. The engineering requirements are related to future applications providing satellite-to-earth connectivity, key network functions to ensure Quality of Service (QoS) in STN. AI has the potential to ensure a threshold of QoS requirements for users, this entails the further development of aggregation algorithms able to tune the network metrics. Breakthroughs involving the use of embedded AI/ML in satellite communications are noted with the goal of increasing the availability and QoS of current and next generation AI space processors [28]. Since satellite mobility is a major challenge in satellite networks, a swarm and location-based approach has been proposed, addressing the routing efficiency with the aim of increasing the QoS [29]. Moreover, a Deep Reinforcement Learning-based resource allocation mechanism for satellite-based Internet of Things is proposed to ease and promote system efficiency, saving costly on-board computing resources [30]. Thus, the trend towards agile and efficient network management and control with the application of SDN has become increasingly important [31].

### 5.3 Connectivity Optimization

The process of enhancing the performance and efficiency of communication networks by improving the connection between various nodes within the network is known as connectivity optimization in both satellite and terrestrial communications. Providing higher data-rate wireless services with global coverage is essential to achieving high connectivity. Emerging approaches based on Green and Red AI bring a focus on efficiency and reduction to gain accuracy. With 6G systems, broad artificial intelligence operations can be implemented to bridge the inherent limitations and challenges of 5G [32]. Pairing satellites with appropriate base stations while channelizing the satellites is a connectivity challenge that has been addressed using Deep Q-Learning approaches [33].

### 5.4 Traffic Management

The global data traffic has increased considerably during the last few years. In addition, it is improbable that traffic explosion will be sustained by the data rate supported by 5G networks [34]. Ensuring efficient management of network resources and effective transmission between network nodes is essentially a matter of controlling and regulating the data flow and information across the network. Moving to 6G communications has significantly improved traffic forecasting, with special emphasis on IoT, autonomous vehicles, smart cities, and telemedicine [35]. A particular issue in traffic management is classifying Internet communications. This can be addressed using a combination of ML and DL techniques to achieve high classification performance [36–39].

Hereafter, Table 1 presents a comparison of the previous technical needs through generations.

**Table 1.** Technical performance comparison table Integrating terrestrial and satellite networks from 3G to 6G.

Aspect	3G	4G	5G	6G
QoS	-Basic support -Limited bandwidth -Moderate latency	-Enhance support -Higher bandwidth -Lower latency	-Network Slicing -Ultra-reliable and low latency	-AI-driven -Dynamic -Multi-dimensional
Network Optimization	-Limited capabilities -Fixed configurations	-Optimization capabilities -Dynamic configurations	-Network slicing -Software defined Networking	-AI-based optimization and adaptation -Network intelligence and optimization
Traffic Management	-Basic prioritization -Quality-based shaping	-Packet-switched -Policy-based	-Software-defined networking -Traffic engineering and optimization	-Smart optimization
Connectivity	-Limited coverage and capacity -Circuit-switched architecture	-Improved coverage and capacity -Packet-switched architecture	-Enhanced mobility support -Efficient handover between cells	-Seamless mobility and handover -Multi-access edge computing

## 6 Architecture Services for IoT-Driven STN

Thanks to tiny, cheap, and low-powered devices performing over long periods with low-cost batteries and high-precision sensors, a growing trend towards the implementation of satellite-based technologies in the short-term and medium-term horizon is highly anticipated. In addition, with global availability of cellular coverage and the widespread use of connected devices across a broad range of industries, this is likely to grant continuous connectivity along with optimized evaluation metrics of the overall communication performance. Distributed computing covers a number of technologies, in this context: cloud, edge, mist and fog computing. They allow a large number of processes to be optimized, with the aim of cutting latency, shortening synchronization times and enhancing node resilience, thus ensuring a globally efficient communication network. Even so, 6G, coupled with these technologies, will undoubtedly address the drawbacks of 5G [40]. Massive resource demand has become a growing concern, which can only be overcome by user-centric approaches [41].

### 6.1 Cloud Computing

Renowned for its high-performance quality, thanks to servers with high storage capacity, cloud computing provides significant calculation performance assets. In addition, with the advent of technological advances in ICT, these servers provide substantial data connectivity and real-time information analysis services. As computing resources are immensely powerful in the cloud, calculations can be off-loaded to it, taking advantage of its unlimited storage resources too [42].

In terms of terrestrial and satellite communications, compliance with 6G is projected over 5G to be better suited to cloud-based services, due to several reasons:

- Enhanced bandwidth and capacity: 6G is set to deliver significantly higher data rates and capacity than 5G. Such increased bandwidth provides faster, increasingly efficient communication between IoT devices and cloud services. Large volumes of data transmission become seamless, thus improving real-time processing, data analysis, and cloud-based resource-intensive applications [43].
- Depressed network latency: with 6G, latency is to be reduced still further compared with 5G. For cloud services, lower latency means real-time interactions and swift response times. In addition, IoT devices with reduced latency are able to reach cloud servers faster, speeding up data decision-making as well as business processing [44].
- Extended network splitting: In 5G, the concept of network splitting has been introduced, enabling the segmentation of a network into multiple virtual networks dedicated to a series of applications. Within 6G, this concept is likely to enable efficient allocation of network resources for cloud-based services. Moreover, this scalability guarantees cloud-based IoT applications get adequate bandwidth, latency and computing strength, thus enhancing all-inclusive system performance [45].

### 6.2 Edge Computing

In the cloud, a vast pool resources can be used to run IoT applications that are intensive in terms of computation. Furthermore, edge computing has emerged to address cloud

computing's high latency and low bandwidth issues. In the terrestrial network, three edge computing architectures are widely used: Cloudlet, Fog computing, and MEC. Cloudlet aims to bring cloud services from remote data centers closer to end users. A cloudlet is simply a micro data center that provides Wi-Fi access to end users. Cloudlets allow end users to deploy and manage their own virtual machines. Fog computing facilitates the implementation of a cloud computing infrastructure, remote from centralized cloud data centers. Fog computing pertains to a cloud that is both collaborative and distributed in nature. The inclusion of numerous geographically dispersed edge nodes constitutes a prevalent feature.

Device management and network management within the network framework are crucial components for ensuring the effective functioning and maintenance of the network infrastructure. These management practices are essential for monitoring and controlling the various network devices, as well as for optimizing network traffic and identifying and resolving network issues. The implementation of Mobile Edge Computing (MEC) has facilitated the provision of cloud computing capabilities within the Radio domain. By utilizing the capabilities of MEC, a plethora of novel applications can be developed. The provision of services. The utilization of Multi-access Edge Computing (MEC) technology can prove advantageous for service providers.

Edge computing integration in LEO networks has the potential to increase the efficiency of satellite IoT systems. Moreover, for large, complex, time-sensitive and intensive IoT applications, it offers an effective solution [46], scrutinizes LEO Edge Computing (LEC) and puts forth a new system architecture for LEC. The architecture presented features a user plane that comprises three layers and a control plane consisting of two levels. The control plane encompasses a worldwide LEC (logical element controller) and controlling agents located on all satellites. The proposed architecture enables the pooling of resources for each satellite, thereby facilitating the efficient utilization of the entire system's resources to process Internet of Things (IoT) data.

The emergence of Satellite-Terrestrial Edge Computing Networks (STECNs) as a widespread solution to facilitate the functioning of multiple Internet of Things (IoT) applications in the context of 6G networks is a subject of current interest in academic circles. The realization of the vision to slice Software-Defined Translucent Electronic Communication Networks (STECNs), discussed in [47] highly depends on the employment of advanced technologies such as Software-Defined Networking (SDN), Network Function Virtualization (NFV), and satellite edge computing, which function as enablers. The establishment of resilient network slicing is imperative for indispensable and vital services. The occurrence of network failures in expansive networks is an inescapable phenomenon that poses a considerable threat to the intricately designed network-slicing infrastructure, thus jeopardizing numerous services. To this end, the authors introduce a robust network-slicing approach that can effectively accommodate malfunctions and ensure uninterrupted service provision. This approach is independent of the integration architecture and possesses innate multi-domain capabilities. Moreover, the authors put forth techniques that aim to attain resilient networking and slicing in Software-Defined



Tactical Edge Communication Networks (STECNs). These encompass the careful organization and allocation of duplicate network resources, formulation of design principles that ensure Service Level Agreement (SLA) breakdown, and implementation of cross-domain approaches for recognizing and counteracting system malfunctions.

To relieve the impediments of the conventional “bent-pipe” satellite constellation, Satellite Edge Computing proposes to position processing resources within LEO constellations [48]. Current work focuses on space-air-ground models supporting the SEC systems. Past these works, authors explore how to effectively implement SEC center-based services to achieve robust coverage with limited resources. Utilizing Lyapunov optimization and Gibbs sampling, with a theoretical performance guarantee, an algorithm for online service placement has been proposed to solve the problems of spatial-temporal dynamic of service coverage and the robustness conflict.

### 6.3 Mist Computing

6G has changed the communication paradigm to a ubiquitous logic supporting the cooperation of a variety of intelligent devices, allowing on-demand services to be delivered from the bottom to the top of the network [49]. The ability of end edge nodes to perform highly complex operations is a major focus of mist computing. This ensures that information can be reliably communicated across the network at low latency and shared in full confidentiality [50]. Mist computing has the potential as an enhancement to satellite and terrestrial communications in 5G and 6G networks, benefiting from reduced computing costs, grid computing capabilities and closeness to the data source. Engaging with these networks might result in:

- Mist computing allows latency reduction by bringing calculations closer to the data source, thus reducing the necessity of transmitting data to distant cloud servers. As this proximity is suitable for STN, it also eases faster response times. Time-sensitive IoT devices can benefit significantly from Mist computing’s low-latency features, thus ensuring a better end-user experience and efficient communications [51].
- Mist computing runs distributed and decentralized, strengthening the resilience and reliability of 5G and 6G network communications. For satellite communications, Mist Computing brings location-based computing power, ensuring non-stop operations at times of disrupted or limited satellite connectivity. Similarly, in terrestrial communications, Mist Computing sets up a reliable network structure involving the distribution of processing resources to various device cores. The resulting network remains uninterrupted, despite node and link failures or congestion [50].
- Mist computing enhances data confidentiality and security in satellite and terrestrial communications. Processing and analysis of sensitive data can be carried out locally. By carrying out these operations at the edge or in fog layers, risk assessment of data security breaches while data is being transmitted to the cloud or satellite is minimized. Furthermore, on-site operations permit successful implementation of data encryption, access control and confidentiality techniques. The integrity and confidentiality of communication are thus assured [52].
- Mist computing ensures scalability and adaptability for accommodating rising numbers of connected devices and data in 5G and 6G networks. Given the ongoing expansion of connected devices, mist computing simplifies computing resource allocation

to support the growing demands of data processing and analysis. Being scalable ensures efficient asset management of the amount and quality of data generated by satellite and terrestrial devices, which in turn permits trouble-free communication within the network architecture [53].

#### **6.4 Blockchain**

Blockchain is a distributed ledger technology in which nodes belonging to the peer-to-peer network architecture are recording each transaction via cryptography-based security approaches. 6G networks are expected to handle both secured and dissimilar devices while ensuring resource efficiency in terms of communication. This focus, however, is constrained by several trust-related issues often left unaddressed in network layouts [54]. In today's mobile communication environment, excessive resource consumption and high latency levels continue to raise challenges that blockchain can potentially address. As such, multiple Consensus Mechanisms have been designed to integrate Blockchain solutions. A prospective analysis of blockchain use in resource sharing within a 6G environment is presented in [55]. Based on the results of this study, several open problems were highlighted, mainly related to IoT devices with low cost, decentralization with high performance, security, and privacy concerns, as well as the assessment of the performance of basic limitations and security issues.

In [56], a wireless access network concept referred to as B-RAN is designed to target emerging wireless communications in which the resources and functionalities are driven by a blockchain framework. Based on the deployment of Fast Smart Contracts (FSCD) to track time-up and secure service delivery while lowering and reducing the risk of malicious or fraudulent requests [57].

### **7 Discussion**

In Core management, AI's contribution to 5G networks is extremely limited in terms of applicability. As for 6G, offering autonomous, integrated AI-based services with all the associated technologies [40]. Consistent decision support is expected to be widely deployed in 6g. Optimal allocation of resources including node clustering and matching via unsupervised learning approaches is achievable [41]. Artificial intelligence, machine learning and deep learning technologies designed for sixth generation (6G) systems are potentially the key to address many technical challenges. Due to its tight QoS requirements, merging these technologies will certainly trigger new services to users and improve the overall performance of terrestrial, satellite and hybrid communications solutions.

For high-data-rate systems, Communication is a key challenge that allows devices to exchange data and information based on network connectivity. From ultra-power hardware design to cloud-based solutions, the discussed IoT communication technologies have steadily evolved representing diverse communication specifications. Based on terrestrial communication networks, these technologies ease conventional types of data communication, using dedicated channels within a local geographical area, and have as their main advantage very low cost. Because of the heterogeneity of technology and

interconnectivity, several emerging IoT domains entail re-modeling, design, and standardization to achieve seamless IoT ecosystems. In fact, wide regions of the globe are still not served by terrestrial networks, leaving behind the IoT global network anytime anywhere principle and the ultra-reliable communication goal. Furthermore, terrestrial communication networks can be damaged and hence connectivity can be lost in the case of natural disasters like earthquakes, tsunamis, floods, volcanic eruption, etc....

Integrating IoT devices with Cloud, Fog and Mist computing on satellite and terrestrial networks presents significant downsides. Foremost, cloud computing relies heavily on centralized hubs, potentially triggering latency and bandwidth restrictions as IoT devices become geographically dispersed, in particular in distant areas and those connected via satellite. Furthermore, although fog computing enables computing to be carried closer to the edge, resource constraints and limited scalability continue to affect its performance. Data confidentiality, data security and reliability may also be an issue given the distributed nature of fog computing resources.

Accordingly, Blockchain may provide key solutions to overcome these drawbacks. Benefitting from the decentralized and secure nature of blockchain, several challenges can be addressed. It ensures transparency and reliability by offering distributed decision-making systems, thereby bypassing centralized governance structures while providing enhanced data exchange security. Thanks to its incorruptible aspect, information integrity is ensured, and falsification is effectively prevented, further strengthening overall system security. In addition, in the cloud, fog and mist across distributed IoT devices, blockchain facilitates efficient resource management and allocation, through optimized use of IoT resources. Taken together, Blockchain technology potentially supplies a highly attractive platform to overcome the IoT deployments in satellite and terrestrial networks.

## 8 Conclusion

Following the growth of IoT, Satellite-Terrestrial networks are gaining importance as next-generation communication systems. In this paper, we proposed a review on terrestrial communication protocols, then we presented open opportunities and challenges regarding the integration of SN and TN. The combination of satellite and terrestrial networks, together with the usage of artificial intelligence (AI), has the potential to completely transform IoT communication. This integration offers seamless and ubiquitous communication for IoT devices by combining the extensive coverage and global connectivity of satellite networks with the efficiency and low latency of terrestrial networks. Overall, the integration of AI and Blockchain into the 6G architecture enhances the fusion of satellite and terrestrial networks, leading to improved resource allocation, security, and edge computing capabilities. This convergence enables seamless IoT connectivity, effective resource allocation, heightened security, and real-time decision-making, driving disruptive IoT applications across industries and propelling the IoT ecosystem to new heights of innovation and connectivity.

## References

1. Hallberg, J., Nilsson, M., Synnes, K.: Positioning with Bluetooth. In: 10th International Conference on Telecommunications, vol. 2, pp. 954–958 (2003)

2. Bisdikian, C.: An overview of the Bluetooth wireless technology. *IEEE Commun. Mag.* **39**(12), 86–94 (2001)
3. Dragomir, D., Gheorghe, L., Costea, S., Radovici, A.: A Survey on Secure Communication Protocols for IoT Systems in International Workshop on Secure Internet of Things (SIoT), pp. 47–62 (2016)
4. Gupta, C., Varshney, G.: An improved authentication scheme for BLE devices with no I/O capabilities. *Comput. Commun.* **200**, 42–53 (2023)
5. Luo, R.: Literature Survey on the Performance of the ZigBee Standard. EE 359 Final Project (2015)
6. Khattak, S.B.A., Nasralla, M.M., Farman, H., Choudhury, N.: Performance Evaluation of an IEEE 802.15. 4-Based Thread Network for Efficient Internet of Things Communications in Smart Cities. *Applied Sciences* **13**(13), 7745 (2023)
7. Marksteiner, S., Exposito Jimenez, V.J., Valiant, H., Zeiner, H.: An overview of wireless IoT protocol security in the smart home domain in *Internet of Things Business Models, Users, and Networks*, pp. 1–8 (2017)
8. Salman, T., Jain, R.: Networking Protocols and Standards for Internet of Things in *Internet of Things and Data Analytics Handbook*, 215–238 (2016)
9. Yassein, M.B., Mardini, W., Khalil, A.: Smart homes automation using Z-wave protocol in *International Conference on Engineering & MIS*, pp. 1–6 (2016)
10. Coskun, V., Ozdenizci, B., Ok, K.: A Survey on Near Field Communication (NFC) Technology in *Wireless Personal Communications* **71**, 2259–2294 (2013)
11. Raso, E., Bianco, G.M., Bracciale, L., Marrocco, G., Occhiuzzi, C., Loreti, P.: Privacy-aware architectures for NFC and RFID sensors in healthcare applications. *Sensors* **22**(24), 9692 (2022)
12. Mulligan, G.: The 6LoWPAN architecture in *Proceedings of the 4th Workshop on Embedded Networked Sensors* (2007)
13. Ee, G.K., Ng, C.K., Noordin, N.K., Ali, B.M.: A Review of 6LoWPAN Routing Protocols in *Proceedings of the Asia-Pacific Advanced Network* **30**(0), 71 (2010)
14. Cameron, N.: LoRa and Microsatellites. In: *ESP32 Formats and Communication: Application of Communication Protocols with ESP32 Microcontroller*, pp. 267–291. Apress, Berkeley, CA (2023)
15. LR Alliance: A technical overview of LoRa and LoRaWAN in white paper, November (2015)
16. Frantz, L.T., Carley, K.M.: A Formal Characterization of Cellular Networks in *SSRN Electronic Journal* (2005)
17. Al-Sarawi, S., Anbar, M., Alieyan, K., Alzubaidi, M.: Internet of Things (IoT) communication protocols: Review in *8th International Conference on Information Technology*, pp. 685–690 (2017)
18. Naik, N.: Choice of effective messaging protocols for IoT systems: MQTT, CoAP, AMQP and HTTP in *IEEE International Systems Engineering Symposium*, pp. 1–7 (2017)
19. Da Cruz, M.A.A., et al.: A Proposal for bridging application layer protocols to HTTP on IoT solutions in *Future Generation Computer Systems* (2019)
20. Ferraz Junior, N., Silva, A.A., Guelfi, A.E., Kofuji, S.T.: Performance evaluation of publish-subscribe systems in IoT using energy-efficient and context-aware secure messages. *Journal of Cloud Computing* **11**(1), 1–17 (2022)
21. Ramyasri, G., Murthy, G.R., Itapu, S., Krishna, S.M.: Data transmission using secure hybrid techniques for smart energy metering devices. *e-Prime-Advances in Electrical Engineering. Electronics and Energy* **4**, 100134 (2023)
22. Saint-Andre, P.: Streaming XML with Jabber/XMPP in *IEEE Internet Computing* **9**(5), 82–89 (Sept.-Oct. 2005)
23. Cao, X., Li, Y., Xiong, X., Wang, J.: Dynamic routings in satellite networks: an overview. *Sensors* **22**(12), 4552 (2022)

24. Jiang, W.: Software defined satellite networks: A survey. *Digital Communications and Networks* (2023)
25. Kota, S., Giambene, G.: 6G integrated non-terrestrial networks: Emerging technologies and challenges. In: *I IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6 (2021)
26. Fang, X., Feng, W., Wei, T., Chen, Y., Ge, N., Wang, C.X.: 5G embraces satellites for 6G ubiquitous IoT: Basic models for integrated satellite terrestrial networks. *IEEE Internet Things J.* **8**(18), 14399–14417 (2021)
27. Niephaus, C., Kretschmer, M., Ghinea, G.: QoS provisioning in converged satellite and terrestrial networks: A survey of the state-of-the-art. *IEEE Comm. Surv. Tutor.* **18**(4), 2415–2441 (2016)
28. Ortiz, F., et al.: Onboard processing in satellite communications using AI accelerators. *Aerospace* **10**(2), 101 (2023)
29. Yang, M., Shao, X., Xue, G., Dou, Y.: Swarm and Location-Based QoS Routing Algorithm in MEO/LEO Double-Layered Satellite Networks. *Wireless Communications and Mobile Computing* (2023)
30. Tang, S., Pan, Z., Hu, G., Wu, Y., Li, Y.: Deep reinforcement learning-based resource allocation for satellite internet of things with diverse QoS guarantee. *Sensors* **22**(8), 2979 (2022)
31. Yuan, S., Peng, M., Sun, Y., Liu, X.: Software defined intelligent satellite-terrestrial integrated networks: Insights and challenges. *Digital Communications and Networks* (2022)
32. Kim, H., Ben-Othman, J., Mokdad, L.: Intelligent Terrestrial and Non-Terrestrial Vehicular Networks with Green AI and Red AI Perspectives. *Sensors* **23**(2), 806 (2023)
33. Yin, Y., et al.: Deep Reinforcement Learning-Based Joint Satellite Scheduling and Resource Allocation in Satellite-Terrestrial Integrated Networks. *Wireless Communications and Mobile Computing* (2022)
34. Alsbah, M., et al.: 6G wireless communications networks: a comprehensive survey. *IEEE Access* **9**, 148191–148243 (2021)
35. Akhtar, M.W., Hassan, S.A., Ghaffar, R., Jung, H., Garg, S., Hossain, M.S.: The shift to 6G communications: vision and requirements. *HCIS* **10**, 1–27 (2020)
36. Pacheco, F., Exposito, E., Gineste, M.: A framework to classify heterogeneous internet traffic with machine learning and deep learning techniques for satellite communications. *Comput. Netw.* **173**, 107213 (2020)
37. Jagannath, A., Jagannath, J., Melodia, T.: Redefining wireless communication for 6G: signal processing meets deep learning with deep unfolding. *IEEE Trans. Artif. Intell.* **2**(6), 528–536 (2021)
38. Ahammed, T.B., Patgiri, R., Nayak, S.: A vision on the artificial intelligence for 6G communication. *ICT Express* (2022)
39. Qadir, Z., Le, K.N., Saeed, N., Munawar, H.S.: Towards 6G internet of things: Recent advances, use cases, and open challenges. *ICT Express* (2022)
40. Banafaa, M., et al.: 6G mobile communication technology: Requirements, targets, applications, challenges, advantages, and opportunities. *Alexandria Engineering Journal* (2022)
41. Li, Q., et al.: 6G cloud-native system: Vision, challenges, architecture framework and enabling technologies. *IEEE Access* **10**, 96602–96625 (2022)
42. Tang, Q., Fei, Z., Li, B., Han, Z.: Computation offloading in LEO satellite networks with hybrid cloud and edge computing. *IEEE Internet Things J.* **8**(11), 9164–9176 (2021)
43. Yang, P., Xiao, Y., Xiao, M., Li, S.: 6G wireless communications: Vision and potential techniques. *IEEE Network* **33**(4), 70–75 (2019)
44. Li, Q., et al.: 6G cloud-native system: Vision, challenges, architecture framework and enabling technologies. *IEEE Access* **10**, 96602–96625 (2022)
45. Foukas, X., Patounas, G., Elmokashfi, A., Marina, M.K.: Network slicing in 5G: Survey and challenges. *IEEE Commun. Mag.* **55**(5), 94–100 (2017)

46. Li, C., Zhang, Y., Xie, R., Hao, X., Huang, T.: Integrating edge computing into low earth orbit satellite networks: architecture and prototype in *IEEE Access* **9**, 39126–39137 (2021)
47. Esmat, H.H., Lorenzo, B., Shi, W.: Towards Resilient Network Slicing for Satellite-Terrestrial Edge Computing IoT in *IEEE Internet of Things Journal*
48. Li, Q., et al.: Service Coverage for Satellite Edge Computing in *IEEE Internet of Things Journal* **9**(1), 695–705 (2022)
49. You, X., et al.: Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts. *Science China Information Sciences* **64**, 1–74, (2021)
50. López Escobar, J.J., Díaz Redondo, R.P., Gil-Castiñeira, F.: In-depth analysis and open challenges of Mist Computing. *Journal of Cloud Computing* **11**(1), 81 (2022)
51. Fazel, E., Najafabadi, H.E., Rezaei, M., Leung, H.: Unlocking the Power of Mist Computing through Clustering Techniques in IoT Networks. *Internet of Things*, 100710 (2023)
52. Ketu, S., Mishra, P.K.: Cloud, fog and mist computing in IoT: an indication of emerging opportunities. *IETE Tech. Rev.* **39**(3), 713–724 (2022)
53. Babar, K., Shah, M.A.: Scalable and sustainable mist computing-based architecture for Internet of Health Things. In: *Competitive Advantage in the Digital Economy (CADE 2022)*, Vol. 2022, pp. 111–116. *IET* (2022 June)
54. Wang, J., Ling, X., Le, Y., Huang, Y., You, X.: Blockchain-enabled wireless communications: a new paradigm towards 6G. *National Science Review* **8**(9), nwab069 (2021)
55. Xu, H., Klaine, P.V., Onireti, O., Cao, B., Imran, M., Zhang, L.: Blockchain-enabled resource management and sharing for 6G communications. *Digital Communications and Networks* **6**(3), 261–269 (2020)
56. Ling, X., Wang, J., Bouchoucha, T., Levy, B.C., Ding, Z.: Blockchain radio access network (B-RAN): Towards decentralized secure radio access paradigm. *IEEE Access* **7**, 9714–9723 (2019)
57. Le, Y., Ling, X., Wang, J., Ding, Z.: Prototype design and test of blockchain radio access network. In: *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6. *IEEE*. (2019)



# BlindEye: Blind Assistance Using Deep Learning

Bilal Shabbir, Ali Salman, Sohaib Akhtar, and M. Asif Naeem<sup>(✉)</sup>

Department of Computer Science, National University of Computer  
and Emerging Sciences (NUCES), Islamabad, Pakistan  
`asif.naeem@nu.edu.pk`

**Abstract.** Blind Community faces significant challenges in independently navigating their surroundings, often relying on white canes or assistance from others. These traditional methods can limit their mobility and autonomy. Therefore, there is a pressing need for innovative solutions that leverage emerging technologies to enhance the navigation capabilities of blind individuals. Deep learning techniques have shown great promise in various computer vision tasks, making them a compelling approach to addressing the complexities of blind navigation. In this study, we are proposing a unified approach that provides blind assistance using obstacle avoidance, terrain awareness and person identification. We utilised U-FCHarDNet model and achieved the state-of-the-art mIOU of 67.82. With a processing time of only 20 ms, our approach is well-suited for real-time environments. By leveraging these advancements, we aim to provide efficient navigation assistance and support to the blind community.

**Keywords:** Blind Assistance · Deep Learning · Pavement detection · Person Identification · Obstacle Avoidance

## 1 Introduction

In recent years, the advancement of technology has brought forth a plethora of smart solutions, leading to significant developments in assistive technologies worldwide. The World Health Organization (WHO) reports that a staggering 253 million individuals worldwide grapple with serious vision impairments, of which 36 million are blind [2]. Although various approaches have been proposed to aid blind individuals in navigating their surroundings within diverse contexts [12], the field of blind assistive technologies has struggled to keep pace with the rapid pace of innovation. Currently, visually impaired individuals rely on their senses and experience, along with the assistance of guide canes, to detect and avoid potential collisions with both moving and stationary obstacles. However, guide canes often fail to meet the necessary safety standards, as they do not provide a comprehensive perception of obstacles or object types, nor do they offer information about the walking path. Consequently, visually impaired individuals must rely on their prior experiences to react in the event of an unforeseen

collision. Despite witnessing numerous technological breakthroughs, there still exist notable limitations in outdoor blind assistance solutions.

After the rising trend in deep learning, several applications have been ascending for blind people. These applications encompass a wide range of functionalities, such as path localization [10, 13], obstacle detection, and more [9]. Numerous approaches have employed diverse techniques to tackle the challenges faced by visually impaired individuals in various contexts. With that consideration in mind, our aim is to address the need for comprehensive and integrated solutions tailored specifically for outdoor navigation, targeting visually impaired individuals. We propose a personal system that combines pavement segmentation, person finding, and obstacle avoidance with semantic segmentation. By integrating these components, we seek to develop a robust and effective outdoor assistance solution, enhancing the navigational experience for visually impaired individuals. To ensure real-time functionality, we are adopting a more robust model. Moreover, existing implementations in the field of blind assistive technologies have not yet provided a unified approach to ensure the safety of visually impaired individuals while guiding them through terrain awareness. The prevalent use of bounding box-based techniques lacks the necessary information to guarantee safety and lacks real-world applicability. While it performs well and is fast, it falls short of providing sufficient information beyond the coordinates of bounding boxes. To overcome these limitations, we employ the Semantic Segmentation approach to train a unified approach for our project. Our goal is to navigate and assist blind individuals, providing them with awareness of their surroundings, and detecting objects and obstacles through voice feedback. We explore various techniques and models to enable blind individuals to differentiate between walking on the road and the sidewalk, as well as avoiding obstacles. Our research involves training DeepLabv3+ with MobileNet backbone, UNet with EfficientNet backbone, and U-FCHarDNet models on our custom dataset. We validate these models using the ADE-20K validation set, comparing their accuracy and prediction time. Our contributions to this research can be summarized as follows:

1. **Custom Dataset Collection:** We have gathered a specialized dataset specifically tailored for blind person navigation, ensuring its relevance to real-world scenarios.
2. **Robust Model Development:** We have trained a robust deep learning model, U-FCHarDNet, capable of making real-time predictions.
3. **Comprehensive Assistance Features:** Our system incorporates key functionalities such as obstacle avoidance, person identification, and pavement guidance. This comprehensive approach aims to provide a holistic navigation experience to visually impaired individuals.

The subsequent sections of the paper are organized as follows: a comprehensive review of related work, followed by the presentation of our proposed approach, which describes our solution in detail. Subsequently, we provide a thorough analysis and discussion of the experimental results in the Experiments section. Lastly, we present our conclusive findings and future directions in the Conclusion section.



## 2 Related Work

Significant improvements have been made in the field of blind assistance in recent years. Researchers have explored various approaches to address the challenges faced by the blind community, leveraging the power of different advanced algorithms and techniques. This section provides an overview of the existing literature, highlighting key studies and recent developments that have contributed to the field. We start off by going over classical approaches background for blind assistance and their uses, and then dig into significant research and approaches using deep learning that has been used to address particular facets of the problem. We also point out research gaps and limits in the literature, which opens the door for our research's innovative contributions. Our goal is to provide a solid foundation for our deep learning-based unified blind assistance system through this comprehensive review, demonstrating its potential to improve the independence and quality of life of people with visual impairments.

Chung et al. introduced a classical approach which included navigation algorithms like path planning and localization with fault handling in the architecture of the navigation system. They proposed a navigation scheme with range sensors and intelligent navigation components. Although, they proposed a system for indoor navigation, but their classical approach was limited in handling the multiple complex scenarios and lacked outdoor navigation [3]. Tudor et al. introduced a blind navigation system using an ultrasonic system. They used HC-SR04 sensor which utilized the sonar technology to determine the distance of an obstacle upfront. Their system provided a signal when the sensor detected the object to assist the blind person [11]. Aladren et al. combined RGBD camera which provides depth information with the image with computer vision approaches to design a navigation system. They utilised depth information with image intensities for floor segmentation [1]. Narayani et al. introduced an approach that utilized PIR sensor and detected objects between chest and knee levels using different height thresholds. After the object is placed in the targeted area, it provided a digital signal with the help of vibration to assist the blind person [8].

Traditional approaches to blind navigation primarily relied on classical methods and sensor-based systems. However, these approaches often face limitations when operating in complex environments. The emergence of machine learning and artificial intelligence (AI) has spurred the development of novel techniques for addressing these challenges. Researchers and innovators in the field of blind assistance have embraced these advancements, leveraging machine learning and AI to explore more effective solutions. Mukhiddinov et al. proposed an approach which helped to detect objects in the night-time and low-light environment. They first enhanced lighting using an exposure-fusion network and then provided feedback to the user as a speech using a transformers-based encoder-decoder model. They achieved state-of-the-art performance on Low-Light and exDark datasets but their approach required high memory and energy requirements which hindered its real-time application capability [7]. Elsonbaty proposed an ultrasonic sensor-based electronic stick which helped to navigate in outdoor and indoor environments. It utilized sensors vibration and speech production to

provide feedback to the blind person. It required the battery to be charged for the assistance [5]. Kuriakose et al. proposed DeepNavi, a deep learning based approach to detect obstacles. Apart from it, their approach also identified the distance from the object and with convenience [6]. Díaz-Toro et al. proposed a vision-based system detecting close obstacles, floor segmentation and vibration feedback. They used RANSAC and a fusion-based approach to detect the desired objects for blind assistance navigation [4]. Existing blind assistance systems have utilized the mentioned approaches; however, these models suffer from high memory requirements and slow inference times. Additionally, they lack a unified system for known person recognition, pavement assistance, and detecting threat severity posed by obstacles. Furthermore, these approaches face limitations in different environments and struggle to perform effectively with local data.

We present a novel unified approach that leverages U-FCharDNet, trained on a custom dataset, that additionally comprises ADEK20 and COCO Context, captured under diverse conditions. Our approach achieves state-of-the-art performance with a mean Intersection over Union (mIOU) of 67.82 on 17 classes. Furthermore, our system offers assistance in local environments and performs exceptionally well with the two widely recognized datasets. Our unified approach encompasses obstacle detection, distance estimation, known person identification through face recognition, pavement assistance, and real-time voice feedback, providing comprehensive support for individuals with visual impairments.

### 3 Proposed Solution

#### 3.1 Execution Architecture

Heeding the ease of the blind community, the main motivation of our project is to help the blind community to move on the track without any stick or a white cane. Our main focus is to help a blind person to be able to go somewhere by avoiding obstacles, getting to know on what track he is walking, switching between road and walking path safely, and opting to get to know whether the road is rushed with cars, besides being aware of coming obstacles and how far they are, to be prone to analyze and save himself using his cognitive abilities. Focusing on the ease of the user, our project aims to be comprised of the following technical devices:

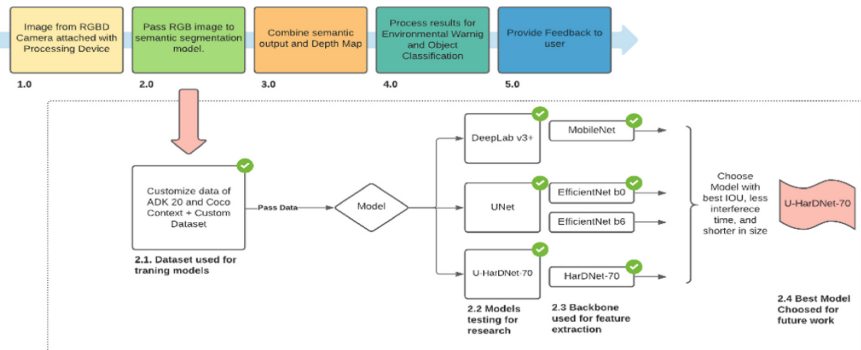
- Processing Device(Raspberry Pi)
- RGB-D sensor for depth perception

The RGB-D sensor will be attached to the processing device capturing the real-time RGB-D streams. The RGB image will be passed to the semantic segmentation model deployed on a local computationally powerful device or a cloud server, while D (Depth Map) is processed to alert the blind person about instant critical obstacles. The output image and depth map are going to be combined and used for environment perception. Perceiving the environment includes classifying objects and their placement with respect to the user’s position. Finally,

the information will be processed to provide meaningful feedback to the user in the form of a voice. Our proposed approach for assisting blind individuals in navigation and obstacle avoidance consists of the following steps:

- Image Capture: An RGB-D camera captures real-time images and depth information of the surroundings.
- Semantic Segmentation: The captured RGB image is processed using a segmentation model to identify objects and classify them accordingly.
- Combination of Semantic Output and Depth Map: The semantic output from the segmentation model and the depth map are combined to gain a comprehensive understanding of the environment, including the placement of objects in relation to the user’s position.
- Post-processing: The combined information is processed to facilitate obstacle avoidance, identify pavement areas, and perform person identification for enhanced safety.
- Voice Feedback: The processed information is relayed to the user in real-time through voice feedback, providing them with meaningful guidance and alerts about the surrounding environment.

Our execution architecture and the end-to-end pipeline are shown in Fig. 1.



**Fig. 1.** After collecting our custom dataset, we trained three architectures and opted for U-FCHarDNet. End-to-End pipeline flow is also shown.

### 3.2 Dataset Selection and Preprocessing

In deep learning models, the availability of appropriate data plays a crucial role. In our research, we utilized two primary datasets: ADK-20 and COCO Context, which contain images with pixel-wise annotations. ADK-20 comprises 20,000 images with 150 different classes, while COCO Context consists of 20,000 images with 180 classes. To ensure the relevance of the data to our specific application and focus on critical classes, we filtered the images, selecting only those that were best suited to our objectives. This filtering process resulted in a total

of 17,370 images from both datasets, along with our custom dataset portion. In addition to the existing datasets, we created a custom dataset by manually labelling 110 images. This addition aimed to enrich the instances of sidewalk annotations, as they were underrepresented in the available data. Apart from it, it enriched the model performance in the local environment as well. Our selected dataset included critical classes like sky, road, grass, sidewalk, person and car. To enhance the diversity and variability of the dataset, we employed data augmentation techniques. Specifically, we applied Color Jitter with brightness set to 0.5, contrast set to 0.5, and saturation set to 0.5. Additionally, we implemented Random Horizontal Flip to generate additional instances and variations within our custom dataset. By carefully selecting and filtering the available datasets, supplementing with our custom annotations, and employing data augmentation techniques, we aimed to create a comprehensive and diverse dataset that adequately represents the classes and scenarios relevant to our research objectives.

### 3.3 Model Selection

One of the most challenging tasks for us was to select a model for semantic segmentation to classify and detect object instances. The semantic segmentation model takes an RGB image as input and returns a semantic image, each pixel in the image represents the label for the respective class. The four main considerations that we had to ruminate upon for selecting any model were as follow:

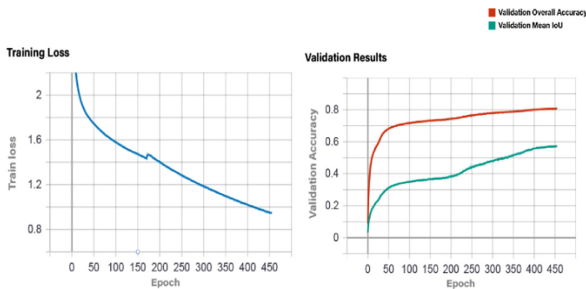
- The model should provide maximum accuracy for real-time application.
- Its size should be small enough to be embedded in a weaker processing device.
- It should not require too much computation power to make predictions (perform single passes).
- It should take less time to predict a result.

After extensive research, we carefully selected three models, each with its own set of advantages and disadvantages. To assess their performance, we trained our data on these models and analyzed the results. Each of these semantic segmentation models consists of two main components: the Encoder and the Decoder. The Encoder incorporates a backbone that extracts features from the entire image, while the Decoder is responsible for transforming these features into the final output, namely the semantic map. Below, we provide a description of the architecture for each of our chosen models.

**DeepLabV3+ with MobileNet:** DeepLabV3+ employed an encoder-decoder technique to convert an image into a semantic image. The encoder module encoded multi-scale contextual information by applying atrous convolution at various scales, while the decoder module refined the segmentation results along object boundaries. DeeplabV3+ offered compatibility with multiple classification models, such as MobileNet, ResNet, Xception, etc., for feature extraction from images. We chose to utilize DeeplabV3+ with the MobileNet backbone for our training process.

**UNet with Efficient Net:** The UNet model followed a U-shaped architecture with an encoder for downsampling to reduce feature map resolution and capture high-level details. In our architecture, EfficientNet was utilized as the backbone, serving as the encoder and enabling the reduction of feature map resolution while capturing important high-level features. The upsampling decoder reconstructed the image from these high-level features to create semantic maps. EfficientNet, the basic building block in our architecture, employed bottleneck convolution (MBCConv). As we progressed from EfficientNet B0 to B7, the model’s depth, width, resolution, and overall size increased, leading to improved accuracy. By incorporating the UNet architecture with EfficientNet as the backbone, our aim was to leverage the advantages of downsampling and upsampling techniques, allowing for precise semantic segmentation and the preservation of fine-grained details throughout the process.

**U-FCHarDNet:** U-FCHarDNet team made modifications to a variant of Dense Net to develop U-FCHarDNet, which offered superior inference times, reducing them by over 30% compared to previous methods. Furthermore, U-FCHarDNet demonstrated faster performance and lower DRAM accesses, resulting in even greater improvements. In fact, there was a notable 40% decrease in DRAM accesses compared to traditional DenseNet approaches. To gain further insights into the traffic flow, ARM Scale Sim and Nvidia Profiler were employed. Following the feature extraction process using the dense net, the upsampling decoder was responsible for reconstructing the image into a semantic map. The training loss graph and the mean IOU of the validation set across each epoch are depicted in Fig. 2.

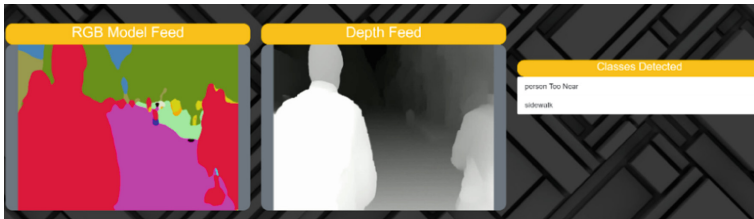


**Fig. 2.** Graphs depicting the training loss & validation accuracy with respect to epochs for U-FCHarDNet

Upon completion of training all models, we selected the model with the highest Intersection over Union (IOU) and Forward Pass Time (FWT) scores. After careful evaluation, we determined that U-FCHardnet exhibited superior performance in terms of mean IOU (mIOU) and FWT. Detailed results can be found in the result section.

### 3.4 Post Processing

In the post-processing stage of our research, we combined the semantic results from the U-FCharDNet model with the depth map obtained from a depth sensor. This fusion enabled us to extract obstacle locations in the environment as the blind person navigating from point A to B. We used Opencv and Python scripts to achieve this combination. To analyze foreground and background objects, we developed a script that categorized the classes into three groups: obstacles, terrain awareness, and person identification. Specifically, we built a Flask web application for demonstration purposes. Within the application pipeline, we initially passed an image from the live feed to the U-FCharDNet model deployed on the Flask server. Subsequently, we implemented a script to combine the semantic map with the depth map, extracting meaningful information such as object class and object distance. We then divided the objects into three main groups: terrain awareness, obstacle avoidance, and person identification. Critical classes such as road, ground, and sidewalk fell under the category of terrain awareness. Voice feedback was provided to the user through the web application, alerting them if they deviated from the road to the sidewalk and vice versa. The web application displayed the model-predicted labels, object positions, and delivered voice feedback as shown in Fig. 3.



**Fig. 3.** Live demonstration of model inference from selected model and Depth map combination with real-time labels prediction

We further classified object positions into three categories: “Too Near,” “Close,” and “Far” (on-demand). Objects within a 2-meter range were labelled as “Too Near,” while “Close” denoted objects located beyond 2m. Additionally, the “Far” label was available on demand, allowing the user to perceive distant objects for improved self-prediction. To ensure efficient voice feedback, our Flask application utilized parallel threads for asynchronous behaviour. We employed the pyttsx3 library, a Python text-to-speech tool, to generate voice feedback. Furthermore, we assigned distinct voice tones to each group, using a female bot’s voice for terrain awareness and a male bot’s voice for obstacle detection. This distinction aimed to enhance comprehension and distinguishability for visually impaired individuals.

## 4 Experiments

### 4.1 Experimental Setup

To ensure efficient and effective training of our models, we employed transfer learning and fine-tuning techniques with weight freezing. All the pre-trained weights from Cityscapes dataset, which resembles our domain. All of our models were trained on a Kaggle GPU for over 50 h. The Adam optimizer, a popular choice in deep learning, was used to optimize the training process. Additionally, normalization techniques were applied to ensure proper scaling and normalization of the data, thereby enhancing the model’s ability to learn meaningful representations. Furthermore, the input images were resized to a standardized size of  $640 \times 480$  pixels, enabling consistent processing across different models. Apart from that, we opted for two loss functions, cross-entropy and Focal loss to optimize hard negative examples.

### 4.2 Evaluation Matrix

As for training a model, it is necessary to set the evaluation criteria, so we will be using the following metrics:

**IOU.** Intuitively, a successful prediction is one that maximizes the overlap between the predicted and the true object. The matrix which we are using to evaluate our prediction is Jaccard Coefficient also known as IoU Matrix. Here,  $A$  and  $B$  are two segmentation masks for a given class (but the formulas are general, that is, you could calculate this for anything, e.g. a circle and a square) and  $\cap$ ,  $\cup$  are the intersection and union operators.

$$Jaccard(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

**FWT.** The time in milliseconds model takes for one forward pass. It will be used to evaluate the speed of our model. It is dependent on factors such as model density in terms of layers, hardware limitations, image resolution etc.

### 4.3 Results and Discussion

In results, Table 1 provides a comparison of mean IOU with various other models, and it is evident that our selected model shows significant improvements in the Road, Sidewalk, Car, and Person classes. These classes are crucial for our application and our contribution aligns with the performance enhancements observed in these specific areas. Also, we achieved state-of-the-art mIOU of 67.82. Moreover, the U-FCHarDNet model demonstrates a lower inference time (Forward pass time or FWT metric) compared to the other models, as presented in Table 2. This improvement can be attributed to its reduced density and

complexity, resulting in superior training performance. In the following sections, we will delve into the detailed results of the models tested on the ADE20k validation dataset.

**Table 1.** Table summarizing IoU across of the 7 critical classes & mean IoU for the various models

Architecture	Sky	Road	Grass	Sidewalk	Person	Car	Water	MeanIoU
Unet	75.3	41.1	52.7	5.3	16.8	21.1	20.4	33.24286
SegNet	91.3	63	62.8	36	31.4	63	58.4	57.98571
Enet	89.7	69.4	56.5	38	26.7	64.8	67.3	58.91429
SQNet	92.2	66.7	65.1	37	31.2	54.2	63	58.48571
LinkNet	91.3	66.3	63.5	35.6	30.6	61	66.6	59.27143
ERFNet	<b>93.2</b>	71.1	64.5	46.1	39.7	70.1	67.9	64.65714
ERF-PSPNet	93	73.8	<b>68.7</b>	51.6	39.4	70.4	<b>77</b>	67.7
Deeplab V3+	89.7	70.3	52.3	52.2	66.9	63.8	45.7	62.98571
<b>U-FCHarDNet</b>	91.7	<b>75.5</b>	52.7	<b>57.9</b>	<b>71.7</b>	<b>70.9</b>	54.4	<b>67.82857</b>

Notably, the IOU for the person class reached 71.7%, demonstrating significant improvement compared to existing state-of-the-art models used for blind navigation assistance. Similarly, the accuracy for the sidewalk class increased by 6% compared to the best-performing model currently available. The forward pass time for a  $640 \times 480$  image was approximately 20 ms, and the model size was 31 MB.

For inference time performance, pertaining to other models we have tested so far, U-FCHarDNet was the best performing one with FWT of 20 ms over our best competitor, ERF-PSPNET, with 34 ms single pass time. These were performed on GPU. The CPU inference time for a lower-end device came out to be 89 ms per single pass. This leads to it performing well in complex scenarios.

**Table 2.** Table summarizing the forward pass times (FWT) for various models

Models	Fwt at $640 \times 480$
Unet	131 ms
SegNet	178 ms
Enet	24 ms
SQNet	89 ms
LinkNet	32 ms
ERFNet	44 ms
ERF-PSPNet	34 ms
Deeplab V3+	26 ms
<b>U-FCHarDNet</b>	<b>20 ms</b>



U-FCHardNet is performing in very accurate manners labelling pink as the colour of sidewalk and pavement, green as the colour of trees and plants, light green as the colour of grass, and red as the colour of a person as shown in Fig. 4. Apart from it, it is labelled sky with blue colour, fence with yellow and building with light blackish grey colour. On the left side is the actual image that is being passed to the model, the middle is the semantic map retrieved from the model and the right side is the semantic map overlayed on the actual image. At the corner of the second and third images, a very little chunk of the sidewalk is given with brown colour mixing the sidewalk with the road and at some pixels, the grass is also mixed with tree and plant labels.

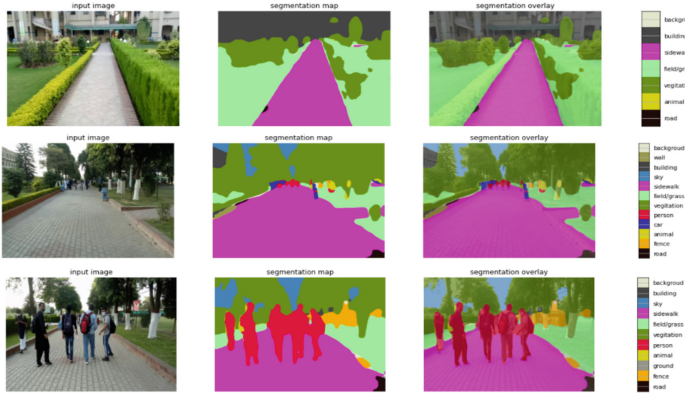


Fig. 4. U-FCHardNet results on different instances in local environment

## 5 Conclusion

In conclusion, our research focuses on assisting blind individuals in outdoor environments by providing them with a sense of freedom through improved terrain awareness and obstacle detection. While the field of terrain awareness for blind assistance is rapidly advancing, the availability of unified models that integrate obstacle detection and terrain awareness remains limited. Through our research, we have achieved notable improvements in accuracy for critical classes, particularly sidewalk and person, which are essential for effective navigation. To achieve these improvements, we selected the FC-Hardnet model, which demonstrated favourable results in terms of reduced forward passing time. However, we also observed lower IoU values for classes such as fence, water, ground, and mountains. Upon investigation, we discovered that these classes had fewer image instances in our training dataset. While adding more images for these classes would enhance their performance, the process is time-consuming, costly, and requires manual labelling. Our research contributes to the development of technologies that enable blind individuals to navigate outdoor environments more

confidently. In future work, we aim to further refine the performance of less critical classes and explore strategies to mitigate the challenges associated with dataset expansion.

## References

1. Aladren, A., López-Nicolás, G., Puig, L., Guerrero, J.J.: Navigation assistance for the visually impaired using RGB-D sensor with range expansion. *IEEE Syst. J.* **10**(3), 922–932 (2014)
2. Bourne, R.R., et al.: Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *Lancet Glob. Health* **5**(9), e888–e897 (2017)
3. Chung, W., Kim, G., Kim, M., Lee, C.: Integrated navigation system for indoor service robots in large-scale environments. In: *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, vol. 5, pp. 5099–5104. IEEE (2004)
4. Díaz-Toro, A.A., Campaña-Bastidas, S.E., Caicedo-Bravo, E.F.: Vision-based system for assisting blind people to wander unknown environments in a safe way. *J. Sens.* **2021**, 1–18 (2021)
5. Elsonbaty, A.A.: Smart blind stick design and implementation. *Int. J. Eng. Adv. Technol. (IJEAT)* **10**(5) (2021)
6. Kuriakose, B., Shrestha, R., Sandnes, F.E.: DeepNAVI: a deep learning based smartphone navigation assistant for people with visual impairments. *Expert Syst. Appl.* **212**, 118720 (2023)
7. Mukhiddinov, M., Cho, J.: Smart glass system using deep learning for the blind and visually impaired. *Electronics* **10**(22), 2756 (2021)
8. Narayani, T.L., Sivapalanirajan, M., Keerthika, B., Ananthi, M., Arunarani, M.: Design of smart cane with integrated camera module for visually impaired people. In: *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, pp. 999–1004. IEEE (2021)
9. Perez-Yus, A., Gutiérrez-Gómez, D., Lopez-Nicolas, G., Guerrero, J.: Stairs detection with odometry-aided traversal from a wearable RGB-D camera. *Comput. Vis. Image Underst.* **154**, 192–205 (2017)
10. Schauerte, B., Koester, D., Martinez, M., Stiefelwagen, R.: Way to go! Detecting open areas ahead of a walking person. In: Agapito, L., Bronstein, M., Rother, C. (eds.) *Computer Vision – ECCV 2014 Workshops. ECCV 2014. LNCS*, vol. 8927, pp. 349–360. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-16199-0\\_25](https://doi.org/10.1007/978-3-319-16199-0_25)
11. Tudor, D., Dobrescu, L., Dobrescu, D.: Ultrasonic electronic system for blind people navigation. In: *2015 E-Health and Bioengineering Conference (EHB)*, pp. 1–4. IEEE (2015)
12. Wang, S., Yu, J.: Everyday information behaviour of the visually impaired in China. *Inf. Res. Int. Electron. J.* **22**(1), n1 (2017)
13. Yang, K., Wang, K., Hu, W., Bai, J.: Expanding the detection of traversable area with realsense for the visually impaired. *Sensors* **16**(11), 1954 (2016)



# Blockchain-Enhanced Privacy and Security in Medicare Data Sharing: Identifying Gaps and Solutions in Current Practices

Abdullah Rehman<sup>1</sup>  and Muhammad Ilyas<sup>2</sup> 

<sup>1</sup> Information Technology, Altinbas University, Dilmenler, Istanbul, 34218 Istanbul, Turkey

abdullah-rehman@outlook.com

<sup>2</sup> Electrical, and Electronics Engineering, Altinbas University, Dilmenler, Istanbul, 34218 Istanbul, Turkey

muhammad.ilyas@altinbas.edu.tr

**Abstract.** Securing Medicare data sharing is of paramount importance. Our current centralized approach to data management falls short of preserving privacy and ensuring safe data sharing. Blockchain technology is explored as a viable solution designed to reinforce security in shared data. We conducted a focused survey to uncover the deficiencies and challenges in existing protocols, which provide valuable insights for the future application of blockchain in patient data exchange. The healthcare sector necessitates immediate data access and entry, yet the risk of data breaches looms large. Our study emphasizes blockchain's potential to enhance data security, granting patients greater control over their shared data. This study is a stepping stone towards a new epoch of data sharing strategies centered on privacy.

**Keywords:** Reliable data sharing · Medicare data · Electronic clinical records · Interoperability · Privacy protection · Centralized data storage · Blockchain architecture · Systematic Literature Review (SLR) · Data exchange · Data storage and management · Data analysis · Blockchain security

## 1 Introduction

The urgency for a reliable exchange of medical information among various stakeholders is increasingly apparent in today's healthcare landscape. The confidentiality of this information necessitates a safe and secure sharing system, an aspect where current technologies fall short due to high maintainability, insufficient throughput, and potential points of failure. As a decentralized architecture, blockchain presents a promising solution with its inherent features such as consensus agreements, transaction ledgers, and hashes, making it a compelling tool for patient healthcare data sharing [1].

Presently, the individual health data or identity verification strategies are inadequately leveraged, leading to fragmented healthcare services. This fragmentation worsens when patients switch doctors or consult multiple healthcare providers. Further complications arise with changes in healthcare coverage, leading to alterations in clinical service providers. Current systems fail to address this issue effectively due to their centralized nature, single point of failure, and difficulty in facilitating information sharing with relevant stakeholders [2].

Effective healthcare management, particularly for chronic conditions, requires the seamless exchange of health information. Disasters that can obliterate existing health data repositories necessitate a robust system that ensures accurate vaccination records, timely patient diagnoses, and prompt evidence for the development and comparative analysis of drugs in the market. Current bottlenecks such as information guidelines and interoperability issues need to be addressed to ensure consistent information exchange and use. Blockchain technology could potentially provide a solution to these challenges, making it a topic of significant interest in healthcare IT discussions [3].

### 1.1 The Objectives of This Study Are

1. The main goal is to create a mechanism for preserving the confidentiality of medical data of patients in Medial Data Block by means of classification of information. This approach will allow better storage and maintenance of information by categories according to its confidentiality and predefined criteria.
2. Another major goal is to develop technology that will improve the privacy of reconfigured information within the Blockchain, with a particular focus on protecting medical data.
3. The goal is to develop a reputation-based approach to evaluating the performance of medical diagnoses, thereby computing the reputation of each diagnosis.

## 2 A Literature Review

### 2.1 Introduction

An urgent research problem in healthcare is the reliable exchange of medical information among various stakeholders [4]. Due to the confidential nature of the data, it needs to be shared securely and reliably. Currently, patient data is stored and shared using centralized methods. These methods, due to high maintenance, low throughput, and single points of failure, fail to optimally achieve data utility while maintaining privacy and confidentiality [9]. Blockchain, with built-in features such as smart contracts, consensus algorithms, and hashing, provides a promising avenue for facilitating integration and achieving the goal of patient data sharing [5].

The concept of a sharing economy is rapidly infiltrating services essential to human well-being, including healthcare. For instance, Teladoc is a telemedicine

platform providing on-demand medical services, with a projected market revenue of \$1.9 to \$30 billion [1]. However, unlike ride-sharing services, sharing in the healthcare space requires organizations to filter patients based on specific conditions, which necessitates access to precise, up-to-date patient data from various healthcare providers [2–4].

Despite this requirement, patients often hesitate to share their medical records due to privacy concerns. In Australia, for instance, over 900,000 individuals opted out of a national health record initiative [6]. Even though 60 percent of healthcare providers recognize the benefits of online patient records [7], there is still significant reluctance, highlighting the need for secure, trustworthy technologies to facilitate health data sharing [8,9].

This reluctance might be attributed to the current centralized approach for storing and sharing health-related data. Centralized systems, such as a doctor's computer or a hospital's file storage, make the process time-consuming, prone to human error, and cumbersome [10]. These systems also fail to provide necessary security, especially during data transfers to third parties, and do not guarantee efficient data exchange and storage [11].

Blockchain technology can resolve these issues. It creates a highly permanent and direct record of exchanges, facilitating easy information exchange and ensuring data protection and security [12]. Furthermore, it mitigates problems associated with modern security mechanisms, such as insider threats, and can help in areas like detecting illegal distribution of expensive drugs [13].

Seven out of ten healthcare leaders, as per a research paper by the IBM Institute, believe that blockchain could revolutionize the healthcare industry, especially in compliance, patient record keeping, and clinical qualification pathways [5]. This thesis attempts to identify current challenges in this space and develop a blockchain-based healthcare framework.

In developing this framework, we specifically address the prerequisites needed for efficient sharing of sensitive patient data. By applying systematic literature review (SLR), we discern research problems in privacy, data classification, and rating calculation. We also explore how smart healthcare systems can play a significant role in addressing security challenges [14,15].

The concept of electronic medical records integrated with cloud services, as proposed in MedShare [16], has shown potential for improving system performance, authentication, and data integrity. The combination of blockchain-based electronic medical records and smart contracts can significantly enhance patient care safety and efficiency [17]. Digital signatures and biometric authentication systems add an extra layer of security for medical records [18]. Along with these, mobile healthcare systems are being considered for storing patient data using blockchain-based electronic medical records [19].

## 2.2 Medical Information Systems

Medical information systems provide online management of patient data. It allows the patient to collect, store, manage and send patient data to the hospital

or wherever else it is needed. Since medical information systems often access, process, or store large amounts of confidential patient data, security is a major concern of these systems [20] as shown in Fig. 1.

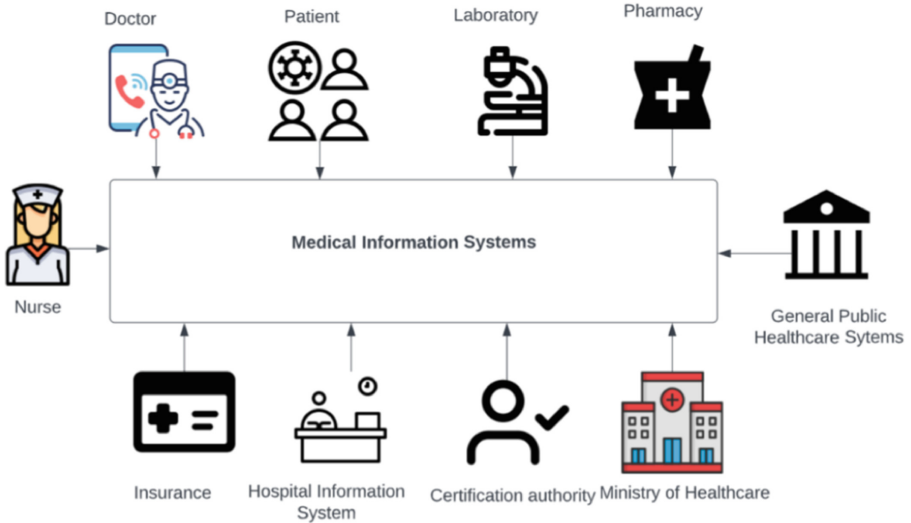


Fig. 1. Caption for your image

### 2.3 Electronic Medical Records

Clinical records can be called electronic health records (EHRs) or electronic clinical records (EMRs), which contain a brief narrative or most recently updated records about a patient,

such as medications, reports, or calorie intake. from a specialist. In an electronic health record (EHR), a patient’s personal medical record can be stored securely and the health care system can be monitored at their convenience [17]. The main function of electronic medical records is the effective management and protection of patients’ medical records [21]. Work with large amounts of clinical data and possibly incorporate complex procedures such as surgery and clinical trials into an established method [22,23]. For a more reflective outcome, health-care data sharing environments facilitate public access to patient healthcare data [24,25]. Information from longitudinal examinations can be used to study the patient’s health throughout his life. Because they contain highly personal and sensitive data, electronic health records and EMRs pose a serious security concern. Patient health data, for example, can be obtained, deleted or changed in the field of electronic health services. Distributed verification of the operation of various system components became possible thanks to blockchain functions [26].

## 2.4 MPI and CDS

The system links patient information between databases using Multi-Patient Information (MPI), reducing duplication of records. It connects individual clinics and reliable cash register systems, using information quality management devices, analysis tools, and access control devices. Patients can check medication and clinical care details with CDS, which helps patients and doctors choose appropriate medical care. Medical care is divided into several parts, with various laws governing healthcare services. HIPAA (Healthcare Diversity and Accountability Act of 1996) and PIPEDA (Privacy and Electronic Records Act) protect healthcare services.

State Health Insurance Law governs hospitals, taking into account wages and taxes. Using the PIPEDA approach, a structured method for medical data transfer is chosen, but consent is required before data can be transferred under HIPAA. GDPR, a European Union privacy and data protection law, allows cross-border data exchange and the free flow of personal data between European countries. Optional uses of Medicare information include electronic medical records, protection cases, and welfare library information.

## 2.5 Medical Information System Security Assessment

One of the significant challenges of medical information systems pertains to the security of IoT devices and the potential for patient data compromise. Malicious attacks can come from various tools, with ransomware and DDoS attacks being particularly prevalent. Ransomware attacks involve hacking into confidential information, potentially leading to misuse of patient data. DDoS attacks can compromise sensitive data, disrupt internet services, and even permanently shut down a healthcare website.

Cybercriminals can interfere with the dissemination of clinical information, exploiting hardware vulnerabilities to encode patient data. They have the capacity to break into the clinical data structure, trace the secret key, and attempt different combinations to penetrate the medical system or a patient's individual system. Such breaches allow attackers to access and potentially transmit patient information to other healthcare systems, and possibly manipulate or confound the patient's clinical data.

## 2.6 Security Requirements

Security and privacy form the cornerstone of clinical data infrastructure, shielding patient information from unauthorized access and interception. Key elements include the Merkel tree method for integration, the collection of vital data through a vertical framework, and granting access exclusively to individuals chosen by the patient. Confidentiality, enhanced by not revealing the patient's secret key to anyone, underpins data security. Crucial is the role of authentication—no one should access a patient's information without consent, controlled via smart contracts that specify access rights. Biometrics also offer

robust verification and security. Interoperability issues, including unauthorized data access or modifications, jeopardize patient well-being and trust in the system. To ensure security and data consistency, smart contracts enforce rules set by patients, and the immutable nature of the registry prevents unauthorized alterations. Despite potential dishonesty in clinical data structures, trust is fostered through patient consent and keys, making unauthorized access to patient information a formidable challenge.

## 2.7 Ethereum and Agreements

Ethereum, launched in 2015, has become a prominent platform for executing smart contracts, primarily facilitating transactions between strangers who may not trust each other due to geographical distance, communication barriers, or existing legal systems' inadequacies. Ethereum uses a transaction ledger and a Turing-Complete scripting language to implement these contracts, ensuring correct transaction execution and state changes.

The consensus protocol, a vital aspect of Blockchain, dictates how nodes add new blocks to the chain. This process works as a predictable lottery system, where each node's chance of winning correlates with its computational power. If a malicious node tries to record lottery results or add a block without accurately executing the agreement, it will be expelled from the Blockchain. This mechanism fosters a system of fair consensus. Such developments have inspired numerous potential uses for Blockchain in various domains, including health information sharing, discussed in subsequent sections.

# 3 Methodology

## 3.1 Research Strategy

This chapter employs the research method known as Design Research (DR) to analyze existing literature and identify challenges in Blockchain's implementation in healthcare. The DR approach involves thoroughly investigating a specific issue, then formulating and evaluating a solution. The process comprises three main stages: characterization of the problem through a literature review, solution structuring, and evaluation. This section covers the first two steps, focusing on creating a Blockchain-based framework for healthcare data sharing.

To conduct the review, the systematic literature review (SLR) procedure is employed, which systematically identifies and analyzes relevant literature in line with the research goal. We complement the basic SLR method with Alzubi et al.'s procedure, which includes defining inclusion and exclusion criteria for paper selection, determining information sources, employing operator data in the selection cycle, and ensuring quality management of selected documents before their analysis.



### 3.2 Inclusion and Exclusion Criteria

In this study, we only include reviews related to Blockchain and its use in medical (or clinical) services. Relevant keywords defined in Table indicating that works related to blockchain, healthcare, Ethereum, or a medical-related field are required to meet the shortlist criteria. This SLR includes targeted research between 2008, when Satoshi Nakamoto published his most memorable article on Bitcoin [27], and 2021. Papers written in English are presented as in this study

**Table 1.** Inclusion and Exclusion Criteria

Category	Keywords
Blockchain	Cryptocurrency, Bitcoin
Health Care	Blockchain, Medical, Permission Less, Authorized
Ethereum	Smart Contracts, Consensus, Hyperledger

### 3.3 Data Collection and Searching Strategy

For the systematic literature review (SLR), we utilized five acclaimed online scientific databases and two search engine platforms. These sources were chosen for their comprehensive coverage of both technical papers and academic studies, encompassing qualitative and quantitative research as well as industry articles. The sources include:

- Elsevier ScienceDirect
- SpringerLink
- Google Scholar
- ACM Digital Library
- IEEE Xplore
- Critical choice of management decision

The search terms from Table 2 were applied to 102 found articles. To be selected for further review, each article had to feature both terms in its headings and keywords. This selection process reduced the count to 44 articles. In the third phase, abstracts of the shortlisted articles were reviewed for relevancy to this study, eliminating those that didn't align with the research objectives. This resulted in a final pool of 36 articles. The rationale behind this phased cycle estimation strategy is summarized in Table 2.

### 3.4 Data Extraction

Following the seven criteria established by Dyb et al. [8] to demonstrate the applicability, reliability and accuracy of the studies, 36 papers were checked for quality assurance at this stage, as shown in Table 2. To ensure the selection of

the most relevant and appropriate articles, we require that an article fit at least four of the paradigms in Table 2. Due to the application of these models, another 24 articles were blocked at this stage, leaving 18 articles, as shown in Figure. These articles have been read for a recent background check and a union to address the intelligence issue in the near future.

**Table 2.** Selection Stages and Corresponding Evaluation

Selection Stage	Applied Method	Evaluation Criteria	Number of Studies
The first stage	Look for the inclusion criteria in the article titles	If the criterion is in the name, accept it, otherwise reject it	102
The second phase	Exclude existing studies on headings and keywords	Title and keywords = search term(s). Yes = Accepted, No = Rejected	44
Third level	Check whether the rest of the research findings are appropriate	If it's appropriate, think about it, otherwise, refuse	36
The fourth stage	Critical analysis of the remaining documents	If other documents discuss or suggest a framework, consider deviating from it	18

## 4 Results Analysis and Interpretation

The emerging realm of healthcare has begun to significantly incorporate the use of digital tools and services. One major avenue in this development is the use of blockchain technologies, specifically in the creation of a novel solution known as Medial Data Block. This system aims to redefine the handling of patient data by focusing on secure storage, privacy, and controlled shareability.

### 4.1 Visual Inspection of Generated Images

The linchpin in the healthcare sector is the safety and security of patients' data. As such, Medial Data Block employs a robust authentication system that exclusively allows verified users to access and utilize data. This is achieved through the implementation of advanced encryption methodologies and hashing algorithms, ensuring the maximum possible security while also maintaining optimal system performance (Zyskind, Nathan & Pentland, 2015)[28].

Integral to the Medial Data Block solution is the integration of a reputation scoring mechanism. This facet of the system allows for legitimate users and healthcare providers to share their experiences and rate their interactions. Consequently, it creates a user-generated content system based on trust, and by

design, remains unaffected by potential manipulation or undue influence (Omar, Bhuiyan, Basu, Kiyomoto & Rahman, 2017) [29].

As an additional layer of protection, Medial Data Block maintains a firm commitment to patient privacy. While it enables data sharing for improving healthcare services, such processes are subject to strict conditions and explicit patient consent. The patients' control over their data is an essential aspect of the Medial Data Block system, and this unique feature sets it apart from many existing solutions in the field.

## 4.2 Data Exchange Through Smart Contracts

At the core of Medial Data Block operations are smart contracts. They provide an automated framework for managing data, handling permissions, and streamlining the consent process (Christidis & Devetsikiotis, 2016) [30]. These contracts predefine the conditions for data access, which can be updated or overridden based on the patient's preferences. This ensures that data exchanges occur strictly within the agreed-upon parameters.

The system uses two central algorithms: one for requesting the owner's data access and permission, and the other for granting access to the requested data. These algorithms rely entirely on the smart contract and a privacy protection mechanism, which keeps the control of health information solely with the data owner (Zhang, Poslad & Ma, 2019) [31].

## 4.3 Interoperability and Compliance

Medial Data Block also addresses a significant challenge in healthcare systems, interoperability. The system is designed to interface seamlessly with existing medical record systems, bridging traditionally isolated information silos, and enabling efficient data sharing among healthcare providers, patients, and potential data users (Kuo, Kim & Ohno-Machado, 2017) [32].

Given the rigorous regulations governing healthcare data, Medial Data Block has been designed to provide comprehensive audit trails ensuring compliance with significant regulations such as HIPAA, GDPR, etc. This makes it a trustworthy system for healthcare providers across different jurisdictions.

## 5 Conclusion and Future Implications

This chapter concludes the current work and outlines avenues for further research. The objective of the research was to delve into the various elements required to create a reliable and feasible off-blockchain healthcare application for health information sharing. This research aimed at developing a comprehensive methodology for creating a system that incorporates all the necessary allocations. Despite the extensive review and analysis of various studies on this topic, there's a significant need for further work on the issue. Future efforts will

expand on several areas related to the thesis, including detailed elaborations on data management, data representation, and data classification.

A key area of future development is the Medial Data Block, which involves more complex and secure Blockchain networks and public network deployment. The Wellbeing Dashboard on Medial Data Block will explore new ways to manage and categorize data on the Blockchain, incorporating various information aggregation and analysis techniques with cloud-based innovations. The introduction of a Reputation Assessment, based on data analysis and hit counts, aims to capture and evaluate user feedback to enhance its usability and web integration potential.

Further, the concept of sub-chains within the blockchain (Medial Data Block) will be developed with privacy features and automatic additions of more child series based on received data. Improving system performance by storing the application in the cloud will also be considered. The Medial Data Block system will also feature smart contract botting, catering to the needs of various internal and external partners, enabling data selection, authorization, and transfer via smart contracts.

## References

1. IBM Watson Health: IBM Watson Health Announces Collaboration to Explore Using Blockchain Technology to Securely Share Health Data. IBM Watson Health, 11 January 2017
2. IBM Watson Health: IBM Watson Health Announces Collaboration to Explore Using Blockchain Technology to Securely Share Health Data. Obtained during a week of clinical trials
3. Tandon, A., Dhir, A., Islam, A.N., Mantemaki, M.: Blockchain in healthcare: a systematic literature review, generalized framework and agenda for future research. *Comput. Ind.* **122**, 103290 (2020)
4. Fusion Practice Survey: Survey: Patients See Average 18.7 Different Doctors. April 27
5. Kitchenham, B.E.A.: Systematic reviews of the software engineering literature - an undergraduate study. *Inf. Technol. Softw.* **52**(8), 792–805 (2010)
6. HIMSS. 2010. 2009
7. Duffy, F.O.: *The Design Research Method* (1999)
8. Eckblau, A., Azaria, J.D.H., Lippman, A.: A case study of blockchain in healthcare: a medrec prototype for electronic medical records and medical research data. In: *Proceedings of the IEEE Open and Big Data Conference* (2016)
9. Ikhsan, M., Jamil, A., Hameed, A.: *Deep learning-based healthcare data analysis system* (2021)
10. Tandon, A., Dhir, A., Islam, A.N., Mantemaki, M.: Blockchain in healthcare: a systematic literature review, generalized framework and agenda for future research. *Comput. Ind.* **122**, 103290 (2020)
11. Al Bawaba: UAE: Using Blockchain Technology to Revolutionize Healthcare in the UAE. Al Bawaba (London) Ltd., 1 June 2016
12. IBM Institute for Business Value: *Blockchains Healthcare to Keep Patients at the Center*. IBM Institute for Business Value Executive Healthcare and Blockchain Report, December 2016

13. Miller, B.J., Moore, D.W., Schmidt, J.C.: Telemedicine and the sharing economy (Year not provided)
14. Practice Fusion: Survey: Patients See Over 18 Different Doctors on Average. Accessed 27 Apr 2010
15. Torre, F., Kotseva, O.R.S., Adorni, G.: The structure of personal data protection in the internet of things. In: 11th International Conference on Internet Technologies and Secure Transactions (ICITST), pp. 384–391. IEEE (2016)
16. Elton, J.: My medical record: more than 900,000 Australians opt out, 18 September 2018
17. Como, J., Ramamoorthy, S., Wallis, J.: Rapid Advances: Rethinking Institutions, Ecosystems, and Economics with Blockchains, New York. Accessed 8 Feb 2017
18. Timitz, K., Lidor, A.: Institute of medicine: crossing the quality hack. In: SAGES Guide to Quality, Outcomes, and Patient Safety, pp. 379–386. Springer (2012)
19. Vest, J.R., Gamm, L.D.: Health information exchange: current challenges and emerging strategies. *J. Am. Med. Inform. Assoc.* **17**(3), 288–294 (2010)
20. Hendry, D.: Three in five GPs report benefit from my medical records: ADHA chief, 20 September 2018
21. Mandel, D., Markwell, R., McDonald, B., Zolovets, E.S.C.: Public standards and patient monitoring: how to keep electronic health records accessible, but confidential. *BMJ* **322**(7281), 283–287 (2001)
22. Smith, P.C.E.A.: Missing clinical information during primary care visits. *Jama* **293**(5), 565–571 (2005)
23. Xia, Q., Sifah, E., Smahi, A., Amofa, S., Zhang, X.: BBDS: blockchain-based data exchange for electronic medical records in cloud environments. *Information* **8**(2), 44 (2017)
24. Xia, Q., Sifah, E., Asamoah, K.O., Gao, J., Du, X., Guizani, M.: MeDshare: trustless medical data sharing among cloud service providers via blockchain. *IEEE Access* **5**, 14757–14767 (2017)
25. Gardner, R.M.: Clinical information systems - from yesterday to tomorrow. *Yearb. Med. Inform.* (Appendix 1), 62 (2016)
26. Roughead, S.A. L., Rosenfeld, E.: Literature Review: Medicine Safety in Australia, Sydney (2013)
27. Esposito, C., De Santis, A., Tortora, G., Chang, H., Choo, K.-K.: Blockchain: a panacea for health-cloud-based data security and privacy? *IEEE Cloud Comput.* **5**(1), 31–37 (2018)
28. Zyskind, G., Nathan, O., Pentland, A.S.: Decentralizing privacy: using blockchain to protect personal data. In: Proceedings of the IEEE Symposium on Security and Privacy Workshops (2015)
29. Omar, I., Bhuiyan, M.Z.A., Basu, A., Kiyomoto, S., Rahman, M.S.: Privacy-friendly platform for healthcare data in cloud based on blockchain environment. *Futur. Gener. Comput. Syst.* **95**, 511–521 (2017)
30. Christidis, K., Devetsikiotis, M.: Blockchains and smart contracts for the internet of things. *IEEE Access* **4**, 2292–2303 (2016)
31. Zhang, P., White, J., Schmidt, D.C., Lenz, G., Rosenbloom, S.T.: FHIRChain: applying blockchain to securely and scalably share clinical data. *Comput. Struct. Biotechnol. J.* **16**, 267–278 (2018)
32. Kuo, T.T., Kim, H.E., Ohno-Machado, L.: Blockchain distributed ledger technologies for biomedical and health care applications. *J. Am. Med. Inform. Assoc.* **24**(6), 1211–1220 (2017)



# Enhanced Image Generation with MorphoGAN: Combining MNNs and GANs

Islam M. Momtaz A. Sadek<sup>1</sup>  and Abdullahi Abdu Ibrahim<sup>2</sup> 

<sup>1</sup> Information Technology, Altinbas University, Dilmenler, Istanbul,  
34218 Istanbul, Turkey

[islam.sadek@alumni.altinbas.edu.tr](mailto:islam.sadek@alumni.altinbas.edu.tr)

<sup>2</sup> Electrical, and Electronics Engineering, Altinbas University, Dilmenler, Istanbul,  
34218 Istanbul, Turkey

[abdullahi.ibrahim@altinbas.edu.tr](mailto:abdullahi.ibrahim@altinbas.edu.tr)

**Abstract.** In this paper, we introduce an innovative technique to enhance the performance of Generative Adversarial Networks (GANs) for image generation tasks by incorporating Morphological Neural Networks (MNNs). The primary objective of this research is to address some of the common limitations associated with GANs, such as mode collapse and training instability, by harnessing the distinctive capabilities of MNNs in analyzing and processing image structures. We put forth a methodology that integrates morphological operations, including dilation and erosion, within the generator and discriminator components of GANs. Our experiments are carried out using the CIFAR-10 data-set, and the performance of our proposed integrated model is compared with multiple established GAN variants. The experimental outcomes reveal that our approach significantly improves image quality, convergence, and stability while maintaining a high level of resistance to noise and artifacts. This study lays the groundwork for further investigation into the synergy of MNNs and GANs across a broad spectrum of image generation applications, presenting valuable avenues for future research.

**Keywords:** Artificial intelligence · Generative adversarial networks (GANs) · Morphological Neural Networks (MNNs) · Conventional Neural Networks (CNN) · Image processing · Performance enhancement · Machine learning · Computer vision · Neural networks · Data creation · Limitations of GANs

## 1 Introduction

Generative Adversarial Networks (GANs), first introduced by Goodfellow [1] have shown remarkable success in generating realistic images, making them a popular choice for various computer vision tasks. Despite their success, GANs

---

A. A. Ibrahim—Contributing author

suffer from several limitations, such as mode collapse, training instability, and sensitivity to noise and artifacts [2]. On the other hand, Morphological Neural Networks (MNNs) are a class of neural networks that incorporate morphological operations, offering unique advantages in processing and understanding image structures [3]. These advantages make MNNs an attractive candidate for integration with GANs to address their limitations.

The main problem addressed in this study is the enhancement of GANs' performance in image generation tasks by integrating MNNs. The integration aims to overcome some of the inherent limitations of GANs and improve their overall performance and stability. This paper proposes a novel methodology for incorporating morphological operations into the generator and discriminator architectures of GANs to achieve this goal. expands on the background of the work (some overlap with the Abstract is acceptable). The introduction should not include subheadings.

### **1.1 The Objectives of This Study Are:**

1. First item To develop a methodology for integrating MNNs into GANs' architectures.
2. To evaluate the performance of the integrated model using various evaluation metrics and comparing it with well-known GAN variants.
3. To demonstrate the practical significance of our proposed method in image generation tasks.

The significance of this study lies in its potential to improve the state-of-the-art GANs' performance, thereby broadening their applicability in various fields, such as medical imaging, art and design, and video game development. Furthermore, this work contributes to the understanding of the synergistic effects of combining GANs and MNNs, opening up new avenues for future research.

### **1.2 The Structure of the Paper is as Follows:**

1. Section 2 provides a review of the literature on GANs and MNNs, highlighting their advantages, limitations, and applications.
2. Section 3 presents the theoretical framework for integrating MNNs with GANs, discussing the motivation, methodology, and potential impacts.
3. Section 4 describes the methodology used in our experiments, including data collection, pre-processing, MNNs design, GAN implementation, and evaluation.
4. Section 5 presents the results, analysis, and interpretation of our experiments.
5. Section 6 concludes the paper, summarizing the main findings, discussing the limitations, and suggesting future directions for research.

## 2 Related Studies on GANS

### 2.1 A Literature Review

Generative Adversarial Networks (GANs), introduced by Goodfellow [1], have become a cornerstone for the generation of realistic data, particularly in the field of computer vision. Despite their success, certain challenges persist, including poor resolution and a restricted range of generated images. Morphology Neural Networks (MNNs) have been proposed to overcome these limitations, aiming to capture the image morphology and enhance the learning capabilities of GANs [2].

One notable advancement in GAN research was the proposal of StyleGAN by Karras [4], a GAN architecture that excelled in generating high-quality, high-resolution images by leveraging adaptive instance normalization (AdaIN) and a novel mapping network. This architecture yielded impressive results on numerous data-sets like CelebA-HQ and FFHQ.

Zhang, addressing the issue of mode collapse, introduced DivCo-GAN [5], a novel architecture that utilizes diversity and competition mechanisms within the discriminator, which improved picture quality and diversity on benchmark data-sets like CIFAR-10 and CelebA.

StarGAN v2 [6], a creation by Choi, presented a unified generative framework that generated diverse cross-domain images by incorporating multi-domain training and an attention mechanism, demonstrating impressive results on various data-sets in terms of image quality and domain adaptation.

Researchers have also explored the potential of GANs for specific tasks. For instance, Lu developed GPT-GAN [7] for text generation, combining the strengths of GPT and GANs to obtain state-of-the-art results on several text generation benchmarks. Also, a GAN-based data augmentation approach, GAN-DA [8], demonstrated improved performance on data-sets like CIFAR-10 and ImageNet.

Furthermore, GANs have been harnessed for image-to-image translation tasks, as demonstrated by Pix2Pix [9] and CycleGAN [10], and for image inpainting tasks by CA-GAN [11]. Also notable is the Super-Resolution Generative Adversarial Network (SRGAN) [12], which produces high-resolution images from low-resolution input, demonstrating exceptional performance in single-image super-resolution tasks.

In summary, the research landscape of GANs has seen significant advancements over the years. The development of novel architectures and training strategies has improved the quality, diversity, and stability of the images generated by GANs, with applications ranging from image-to-image translation to data augmentation. This review aims to offer a comprehensive understanding of the effect of utilizing MNNs to enhance GAN performance in image generation [2]. The lessons learned from these studies will continue to shape the future of GAN research.



## 2.2 Studies on Using MNNs to Enhance the Performance of GANs

Combining Morphological Neural Networks (MNNs) with Generative Adversarial Networks (GANs) can offer potential improvements in the performance of GANs in various image generation tasks. Within this section, we will examine recent studies that investigate the integration of MNNs with GANs to enhance their abilities. By combining these two powerful techniques, researchers hope to leverage the strengths of both models to push the boundaries of image processing and achieve even more impressive results.

Smith and his colleagues (2019) introduced a Morphological GAN (MGAN) architecture designed for texture synthesis. By incorporating morphological operations into the GAN framework, MGAN can better capture the local structures of textures, resulting in improved texture synthesis quality. The 3 authors evaluated their method on multiple texture data-sets and obtained cutting-edge outcomes in terms of the quality and speed of the synthesis [13].

A Morphology GAN (MGAN) architecture was proposed by Lee and his colleagues (2020) for challenges involving image inpainting, which entails repairing damaged or missing portions of images. By using morphological operations in both generator and discriminator networks, The proposed MGAN method can effectively restore the missing portions of images while preserving the structural coherence of the inpainted regions. The authors conducted evaluations of their approach on various data-sets, including the Places2 data-set, and obtained remarkable results, showcasing state-of-the-art performance in both inpainting quality and processing speed [14].

A Morphology GAN (MGAN) for single-image super-resolution tasks was presented by Deng and his colleagues in 2021 [15]. The MGAN incorporates morphological operations in the generator network, enabling it to generate high-resolution images with sharp edges and fine details. The authors evaluated their method on several benchmark data-sets, such as Set5, Set14, and BSD100, and achieved state-of-the-art results in terms of super-resolution quality and speed [15].

Zhou and his team in 2020 developed a Morphological CycleGAN (MCGAN) [16] approach for unpaired image-to-image translation tasks. This integration of morphological operations into the CycleGAN framework improves the ability of MCGAN to preserve the local structures and semantic information of the translated images. The effectiveness of this approach was evaluated on various data-sets, including Maps and Cityscapes, where it delivered outstanding results, setting a new benchmark in translation quality and diversity [16].

## 3 Methodology

In this section, we provide a brief overview of GANs and MNNs, their architectures, and limitations. We also explain the motivation and theoretical framework for integrating MNNs with GANs, and describe the data collection and preprocessing, MNN design, implementation and training of GANs, and model evaluation. GANs and MNNs Overview: Generative Adversarial Networks (GANs) are

a class of deep learning models that consist of two neural networks, a generator and a discriminator, working in tandem to generate realistic data samples [1]. The generator synthesizes new data samples, while the discriminator evaluates the authenticity of the generated samples. Mixed Neural Networks (MNNs) are neural networks that combine different types of layers and activation functions to improve performance and versatility [17].

**Limitations of GANs and MNNs:** Despite their successes, GANs suffer from several challenges, including mode collapse, training instability, and difficulty in generating high-quality images [18]. MNNs, on the other hand, can experience difficulties in optimization and convergence due to the combination of various layer types and activation functions [19].

**Motivation for Integrating MNNs with GANs:** The integration of MNNs with GANs aims to address the limitations of both models by leveraging the strengths of MNNs to improve GAN performance. By incorporating MNNs into the generator and discriminator, the resulting GAN can benefit from increased flexibility, robustness, and expressivity, leading to improved image generation and stability [20].

**Data Collection and Preprocessing:** The data used in this study were collected from publicly available image data-sets, such as CelebA-HQ and FFHQ. Preprocessing steps included resizing images, data normalization, and data augmentation through techniques like random cropping, flipping, and rotation to increase the diversity of the training set.

**MNN Design:** The MNN architecture was designed by combining convolutional layers, fully connected layers, and various activation functions, such as ReLU, Leaky ReLU, and Tanh. The MNNs were incorporated into both the generator and discriminator networks of the GAN.

**Implementation and Training of GANs:** The GANs were implemented using TensorFlow, and the MNN-based generator and discriminator were trained using the Adam optimizer with a learning rate scheduler. The training process included alternating between generator and discriminator updates, with the generator minimizing the difference between generated samples and real samples, while the discriminator aimed to maximize the difference.

**Model Evaluation:** The performance of the MNN-based GAN was evaluated using several metrics, including the Fréchet Inception Distance (FID) and the Inception Score (IS) to assess the quality and diversity of the generated images. Additionally, qualitative visual assessments were conducted to analyze the overall realism and details of the generated images.

## 4 Results Analysis and Interpretation

### 4.1 Visual Inspection of Generated Images

In this sub-section, we present the generated images from the four GAN models, DCGAN, DCGAN w/ MNNs, IMGAN/CNNs, and IMGAN/MNNs, to visually compare the results. This comparison allows us to evaluate the impact of incorporating Morphological Neural Networks (MNNs) into the GAN architecture on

image quality and overall performance. The visual inspection provides a human-centric perspective, which is crucial for applications where the visual appeal of the generated images is of paramount importance [9].

The generated images from each model were produced under the same conditions, utilizing the same data-set and hyper-parameters. A side-by-side comparison of the images generated by the four models is presented below in Fig. 1.

By comparing the images generated by each model, we can observe the impact of incorporating MNNs into the GAN architecture. The comparison highlights the differences in image quality and reveals the potential benefits of using MNNs for enhancing the performance of GANs. This qualitative assessment is a valuable complement to the quantitative metrics, such as FID and IS, discussed in the previous sections.

## 4.2 Quantitative Analysis of Model Performance

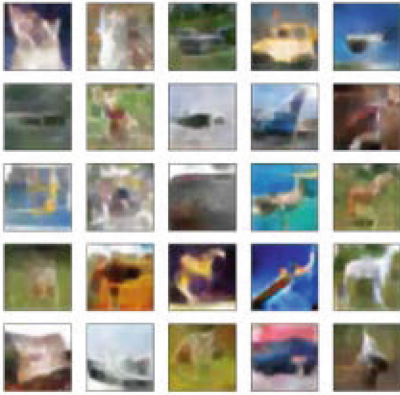
In addition to visual inspection, the Inception Score and Fréchet Inception Distance metrics were used to quantitatively evaluate the performance of the four GAN models. These metrics provided objective measurements of the quality and similarity of the generated images compared to the real images, allowing for a more rigorous and unbiased comparison of the models. By analyzing these metrics, we were able to determine which model(s) excelled in terms of image quality, diversity, and realism [21].

In our experiments, we ensured that the tests made on DCGAN and DCGAN w/ MNNs were conducted in the same environment with the same hyper-parameters. This approach allowed us to make a fair comparison between the baseline DCGAN model and the modified DCGAN with MNNs. By keeping the environment and hyper-parameters constant, we could attribute the observed differences in performance to the incorporation of MNNs in the GAN architecture.

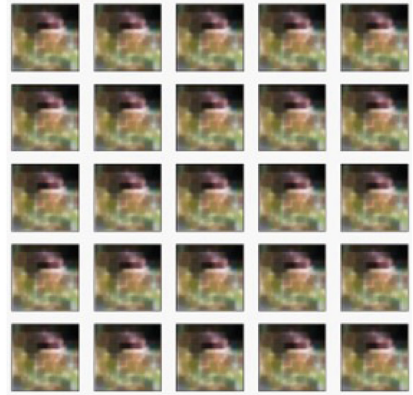
## 4.3 Quantitative Analysis of Model Performance

In addition to visual inspection, the Inception Score and Fréchet Inception Distance metrics were used to quantitatively evaluate the performance of the four GAN models. These metrics provided objective measurements of the quality and similarity of the generated images compared to the real images, allowing for a more rigorous and unbiased comparison of the models. By analyzing these metrics, we were able to determine which model(s) excelled in terms of image quality, diversity, and realism [21].

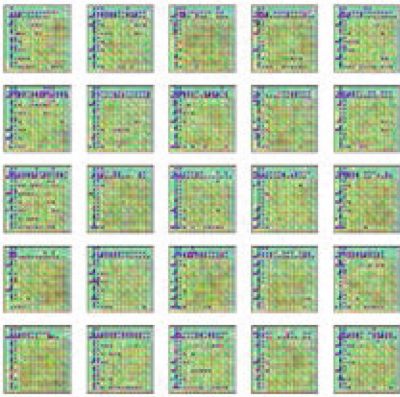
In our experiments, we ensured that the tests made on DCGAN and DCGAN w/ MNNs were conducted in the same environment with the same hyper-parameters. This approach allowed us to make a fair comparison between the baseline DCGAN model and the modified DCGAN with MNNs. By keeping the environment and hyper-parameters constant, we could attribute the observed differences in performance to the incorporation of MNNs in the GAN architecture.



(a) Sample images generated by the IMGAN model with MNNs.



(b) Sample images generated by the IMGAN model with extra CNNs.



(c) Sample images generated by the DCGAN model with MNNs.



(d) Sample images generated by the baseline DCGAN model.

**Fig. 1.** Visual inspection of generated images.. (GAN code is available and can be submitted upon request)

#### 4.4 Impact of Morphological Operations on the GANs

Understanding the effects of incorporating morphological operations (dilation and erosion) in generator and discriminator models is crucial for determining their potential usefulness in GANs for image synthesis. By comparing the performance of the modified GAN models (DCGAN with MNNs, IM-GAN with extra CNNs, and IM-GAN with MNNs) with the baseline Original DCGAN model, we identified the benefits and drawbacks of using morphological operations in the context of GANs. This analysis provided insights into the impact of these operations on generated image quality, diversity, and realism, and helped us uncover any potential advantages or limitations of integrating morphological operations into GAN architectures.

Our experiments indicate that incorporating MNNs and additional CNN layers in the GAN architecture can lead to improved image generation quality, even when using a data set like CIFAR-10, which is not ideal for generating high-resolution images. Visual inspection of the generated images revealed that models with MNNs and additional CNN layers produced more visually coherent and realistic images compared to the Original DCGAN. Furthermore, the quantitative analysis demonstrated that the modified GAN models achieved higher Inception Scores and lower FID values, indicating their enhanced performance.

Through this analysis, we aimed to answer several important research questions, such as whether morphological operations could improve the quality of the generated images, or if they introduced any unwanted artifacts or distortions. Furthermore, we investigated whether the use of morphological operations in the generator and discriminator models had a synergistic effect that could lead to better overall performance.

In this section, we analyze the impact of morphological operations on GAN performance. We incorporated morphological operations into both the generator and discriminator models. The generator model architecture now includes three pairs of dilation and erosion layers to perform morphological closing and opening operations. These layers are added after the second convolutional transpose layer and before the final convolutional transpose layer. The discriminator model has two pairs of dilation and erosion layers to perform morphological closing and opening operations, following the second convolutional layer and before the final dense layer [22, 23].

These morphological operations are expected to enhance the GAN's ability to capture image structures and improve overall performance. Our experiments evaluate the impact of these morphological operations on the Inception Score (IS) and Fréchet Inception Distance (FID) metrics as presented above in Fig. 2 and in Fig. 3.

The morphological operations used in our GAN models are dilation and erosion, Equations 1 and 2 show the mathematical formulas for the dilation and erosion operations, respectively. Equations 3 and 4 shows the generator and discriminator models in Our IMGAN w/ MNNs, respectively. All are defined mathematically as shown in the Figures below

$$(f \oplus g)(x, y) = \max_{(m,n) \in G} \{f(x - m, y - n) + g(m, n)\} \quad (1)$$

$$(f \ominus g)(x, y) = \min_{(m,n) \in G} \{f(x + m, y + n) - g(m, n)\} \quad (2)$$

$$\hat{y} = \tanh(w_3(d_3(e_3(d_2(e_2(d_1(e_1(\text{ReLU}(w_2(\text{BN}(\text{ReLU}(w_1 z))))))))))) \quad (3)$$

$$\hat{y} = \sigma(w_3(d_2(e_2(d_1(e_1(\text{ReLU}(w_2(\text{ReLU}(w_1 x)))))))) \quad (4)$$

**Fig. 2.** Equations for mathematical morphology and deep learning models.

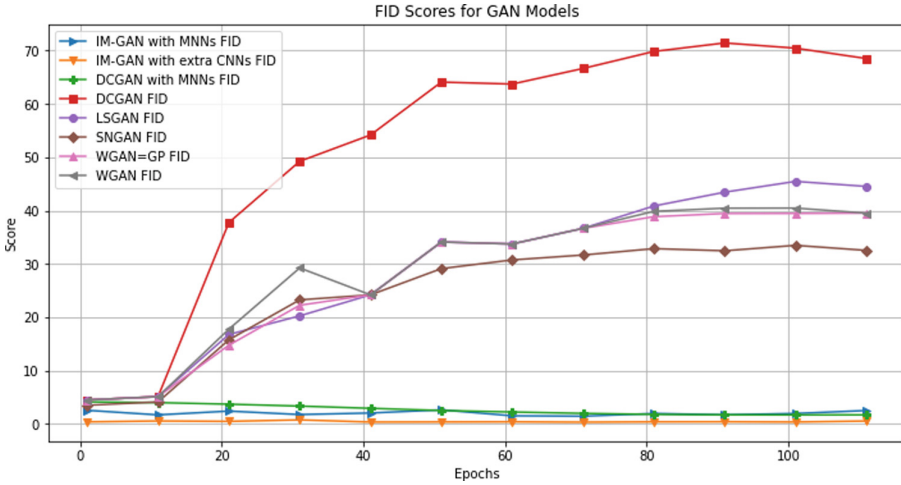
#### 4.5 Comparison of GAN Performance

This section compares our GAN models' performance with MNNs and popular GANs trained on the CIFAR-10 data set. The Inception Scores (IS) and Fréchet Inception Distances (FID) for each model are provided in Table 1 below:

**Table 1.** Comparison of Inception Scores and FID values between some famous GANs, DCGAN, DCGAN with MNNs, IM-GAN with extra CNNs, and IM-GAN with MNNs trained on CIFAR-10 data-set

Model	Epoch	IS	FID
Original DCGAN	121	1.00016	69.68781
DCGAN with MNNs	121	1.00001	1.74414
IM-GAN with extra CNNs	120	1.00048	0.50222
IM-GAN with MNNs	120	1.00001	2.47258
WGAN	–	3.82	40.2
WGAN-GP	–	4.48	39.6
LSGAN	–	3.67	45.9
SNGAN	–	6.35	33.3

Our experiments show an improvement in the FID, indicating that the quality of the generated images has improved. Compared to popular GANs, our results can be considered good in terms of both the Inception Score and the Fréchet Inception Distance. The incorporation of MNNs in the DCGAN and IM-GAN models led to a significant improvement in their performance, thereby supporting our main goal of adding MNNs to enhance the performance of GANs. This comparison serves as evidence of the effectiveness of our proposed modifications to GANs and supports the hypothesis that incorporating MNNs can lead to improved performance in generating high-quality images. In summary, the evaluation and analysis of the four GAN models (Original DCGAN, DCGAN with MNNs, IM-GAN with extra CNNs, and IM-GAN with MNNs) provided valuable insights into the impact of incorporating morphological operations and additional CNN layers into the GAN architecture. This study serves as a foundation for further exploration of the potential benefits and applications of morphological operations in GANs and other generative models.



**Fig. 3.** FID Scores for GAN Models (Drawn by Python Graphic and code is available upon request)

## 5 Conclusion

In our study, we devised and applied four distinct GAN models to explore the impact of adding morphological operations and extra CNN layers into the GAN architecture. Our performance evaluation was based on visual inspection, Inception Score, and Fréchet Inception Distance, allowing us to assess the influence of morphological operations on the generated images and the overall GAN performance. Our findings underline the potential of these operations to enhance the quality of generated images and stress the necessity for careful design and implementation within both generator and discriminator models to maximize performance. This research offers significant insights into the utility of morphological operations in GANs and forms a foundation for future investigations aimed at improving GAN performance and expanding the use of morphological operations in various generative models and tasks.

## 6 Limitations and Future Work

During implementation, we encountered challenges such as increased model complexity, extended training times, and heightened computational resource requirements when integrating morphological operations. Deciding optimal hyperparameters, particularly for morphological operations, proved complex and required extensive trial. Limitations arose from the use of Google Colab for training and the CIFAR-10 data set for high-resolution image generation. Despite these, the promising results encourage further exploration into advanced morphological operations and their impact on varied GAN architectures like StyleGAN or CycleGAN. Assessment of the proposed alterations' efficacy on diverse data

sets and tasks would expand their applicability. Furthermore, we plan to focus on enhancing the Inception Score to generate higher-quality images, thereby boosting the overall performance of our GAN architecture.

## References

1. Goodfellow, I.J., et al.: Generative adversarial networks. arXiv preprint: [arXiv:1406.2661](https://arxiv.org/abs/1406.2661) (2014)
2. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein GAN. arXiv preprint: [arXiv:1701.07875](https://arxiv.org/abs/1701.07875) (2017)
3. Sussner, P., Ritter, H.: An adaptive morphological neural network for image processing tasks. In: Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205), pp. 561–564 (2001)
4. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. arXiv preprint: [arXiv:1812.04948](https://arxiv.org/abs/1812.04948) (2018)
5. Zhang, H., Zhang, Y., Wang, T.: DivCo: diverse conditional image generation with GANs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2867–2876 (2020)
6. Choi, Y., Uh, Y., Yoo, J., Ha, J.-W.: StarGAN v2: diverse image synthesis for multiple domains. arXiv preprint: [arXiv:1912.01865](https://arxiv.org/abs/1912.01865) (2020)
7. Lu, C., Yu, F., Zhan, X.: GPT-GAN: a generative adversarial network for text generation based on the GPT model. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 15097–15104 (2021)
8. Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., Yang, M.-H.: Adain-styleGAN: generating high-resolution images with adaptive instance normalization. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**(9), 3140–3149 (2020)
9. Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv preprint: [arXiv:1703.10593](https://arxiv.org/abs/1703.10593) (2017)
10. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Generative image inpainting with contextual attention. arXiv preprint: [arXiv:1801.07892](https://arxiv.org/abs/1801.07892) (2018)
11. Ledig, C., et. al: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4681–4690 (2017)
12. Ni, J., Schneiderman, H.: An Improved Texture Synthesis Algorithm Using Morphological Processing with Image Analogy. Carnegie Mellon University (2004). Accessed <https://www.ri.cmu.edu/>
13. Chen, J., Xu, Q., Kang, Q., Zhou, M.: MOGAN: Morphologic-Structure-Aware Generative Learning from a Single Image. arXiv preprint [arXiv:2103.02997](https://arxiv.org/abs/2103.02997) (2021). Accessed <https://arxiv.org/abs/2103.02997>
14. Wang, T., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional GANs. arXiv preprint: [arXiv:1711.11585](https://arxiv.org/abs/1711.11585) (2018)
15. Park, T., Liu, M.-Y., Wang, T., Zhu, J.-Y.: Semantic image synthesis with spatially-adaptive normalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2337–2346 (2019)
16. Liu, M.-Y., et al.: Swapping autoencoder for deep image manipulation. arXiv preprint: [arXiv:2007.00653](https://arxiv.org/abs/2007.00653) (2020)



17. Brock, A., Donahue, J., Simonyan, K.: Large scale GAN training for high fidelity natural image synthesis. arXiv preprint: [arXiv:1809.11096](https://arxiv.org/abs/1809.11096) (2019)
18. Lin, Z., Khetan, A., Fanti, G., Oh, S.: PacGAN: The power of two samples in generative adversarial networks. In: Bengio, S., et al. (eds.) *Advances in Neural Information Processing Systems*, vol. 31. Curran Associates, Inc. (2018). Accessed <https://proceedings.neurips.cc/paper/2018/file/288cc0ff022877bd3df94bc9360b9c5d-Paper.pdf>
19. Alotaibi, A.: Deep generative adversarial networks for image-to-image translation: a review. *Symmetry*, **12**(10), 1705 (2020). Accessed <https://www.mdpi.com/2073-8994/12/10/1705>
20. Denton, E., Chintala, S., Szlam, A., Fergus, R.: Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks. [arXiv:1506.05751](https://arxiv.org/abs/1506.05751) [cs.CV] (2015). Accessed <https://doi.org/10.48550/arXiv.1506.05751>
21. Dosovitskiy, A., Brox, T.: Generating Images with Perceptual Similarity Metrics based on Deep Networks. [arXiv:1602.02644](https://arxiv.org/abs/1602.02644) [cs.LG] (2016). Accessed <https://doi.org/10.48550/arXiv.1602.02644>
22. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training GANs. arXiv preprint: [arXiv:1606.03498](https://arxiv.org/abs/1606.03498) (2016)
23. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of GANs for improved quality, stability, and variation. [arXiv:1710.10196](https://arxiv.org/abs/1710.10196) (2017). Accessed <https://arxiv.org/abs/1710.10196>



# Implementation of Homomorphic Encryption Schemes in Fog Computing

Shrayash Pandey<sup>1</sup>(✉), Bharat Bhushan<sup>1</sup>, Alaa Ali Hameed<sup>2</sup>, Akhtar Jamil<sup>3</sup>,  
and Aayush Juyal<sup>1</sup>

<sup>1</sup> Department of Computer Science and Engineering School of Engineering and Technology,  
Sharda University, Greater Noida, India

shrayash.pandey@gmail.com

<sup>2</sup> Department of Computer Engineering, Istinye University, Istanbul, Turkey

alaa.hameed@istinye.edu.tr

<sup>3</sup> Department of Computer Science, National University of Computer and Engineering,  
Islamabad, Pakistan

akhtar.jamil@nu.edu.pk

**Abstract.** In today's date, fog computing is on the rise due to its characteristics. Its ability to provide smart devices with close-by computing capabilities offers a drastic reduction in cloud traffic and more efficient data transfer. However, data security in fog computing devices is a major concern. Homomorphic Encryption (HME) Scheme is a technique that protects private data from various threats, and to improve efficiency and reduce time, it offers modification of original data after encryption without decrypting it. In this paper, we explore the various homomorphic encryption schemes that can be implemented in fog computing to protect the data and provide data security. We present a detailed architecture and various characteristics of fog computing to better visualize the use of homomorphic encryption in different areas. Further, we present the applications of homomorphic encryption in different sectors and discuss future research directions.

**Keywords:** fog computing · homomorphic encryption schemes

## 1 Introduction

Internet of Things (IoT) systems are equipped with loosely connected devices that are interconnected via heterogeneous networks. Such systems consist of two types of data: Little Data or Big Stream, and Big Data [1]. Little Data refers to temporary data which is collected from IoT smart devices at a constant pace. On the other hand, Big Data is permanent data that is kept in centralized cloud storage, and later archived. Different IoT sectors such smart cities and healthcare consist of both, Little Data and Big Data. Decision making tasks, and real-time analytics require both types of data. Mostly, the data is captured from smart devices in the IoT network and then sent to the cloud servers for it to be processed and stored. This process is referred to as Cloud Computing [2]. When the data is stored in the cloud servers, cloud computing fulfills the requirements of Big

Data processing. On-demand scalability of clouds enables processing of huge amounts of data. However, cloud computing data processing doesn't meet the requirements when there's a requirement of low latency requirement and data sources are sent to multiple locations. Therefore, the concept of Fog Computing came into play in 2012, which enabled its services to the edge computing and offered similar computing capabilities of cloud computing with low-latency.

Fog computing provides low latency with the help of its distributed architecture which offers an enriched layer between the cloud and edge devices. The additional layer offers high security and high computing performance. Routers, switches, and virtualized servers are referred to as fog nodes in the network which connect the cloud servers to the edge nodes via internet. Sensors and other devices at the edge generate data, which is subsequently updated by fog nodes and delivered to a storage unit on the cloud. With the advent of cybercrime, it is vital to preserve public data while maintaining privacy. Several private domains provide such amenities; however, they are significantly more expensive than those provided by the public domain. As a result, individuals store and modify the data via the internet using cloud storage. Various encryption techniques have been implemented in this area to safeguard data on the cloud [3]. Data security and protection are crucial as the internet and public clouds are increasingly used for data storage. The information may be encrypted and saved in any database. However, decryption is necessary to execute specific data operations. The data is not secure; therefore, a homomorphic encryption approach is developed to conduct calculations on the cipher text [4]. Users can do additions and multiplications on the initial text using the HME [5].

In the IT industry, data is typically kept and processed on third-party servers such as the cloud, and service providers. As a result, two primary issues arise. Firstly, older encryption systems do not ensure third-party data security. Data may be misused by a third party. Secondly, before performing any actions on encrypted data, it must first be decrypted. These issues lead to the development of homomorphic encryption, which allows conducting computations on encrypted information stored on different servers without decrypting it. The concept of executing computation over data that is encrypted, known as privacy homomorphism [6], first appeared in 1978. RSA was also introduced later in the same year. RSA [6] is an asymmetric encryption system that uses a public key for encryption and a secret key for decryption that is only known to the authorized user. Only multiplication operations on encrypted data are permitted by RSA. Fully homomorphic encryption (FHE) is a HME that permits limitless multiplication and addition operations on encrypted data. FHE remained an unsolved challenge. Many scholars presented strategies, including the El-Gamal scheme [7], and the Paillier system [8]. A summary of the major contribution of this work is mentioned below.

- This work presents various characteristics, a detailed architecture, and different applications of fog computing to identify the required security measures.
- This work provides an in-depth review of HME schemes and analyzes different schemes to protect data within fog computing.
- This work highlights the different sectors in which fog computing plays a major role to implement homomorphic encryption schemes and provide data security.

The remainder of the paper is organized as follows. Section 2 provides an overview of fog computing with a list of its characteristics and architecture. Section 3 explains

homomorphic encryption schemes and the classifications of it. Section 4 explores and analyzes various homomorphic encryption schemes to be implemented in fog computing. Further, Sect. 5 identifies different areas for implementation of fog computing and the data security measures needed. Finally, Sect. 6 presents the conclusions followed by future research directions.

## 2 Fog Computing

Fog computing offers various services which tend to lack in the cloud and edge computing. Therefore, fog computing has an advantage over them due to its dynamic characteristics [9, 10]. Various characteristics that define the properties of fog computing along with the architecture of fog computing are explained and illustrated in the subsections below.

### 2.1 Characteristics of Fog Computing

#### Scalability

Despite having infinite virtual resources, if the end device's data is constantly sent to the cloud, it can turn the cloud into a bottleneck. Fog computing reduces the stress of processing data on the cloud since it processes the incoming data closer to the data source. Therefore, it addresses the issues related to scalability due to the high number of endpoints.

#### Low-Latency Requirement

Antilock brakes on a vehicle are a great example of mission-critical applications. Such applications require data processing in real-time. The feedback of control systems and data obtained by the sensors are the depending factors for motion control in a robot. Fog computing makes real-time response possible by performing the processing close to the robots.

#### Suitable for IoT Tasks and Queries

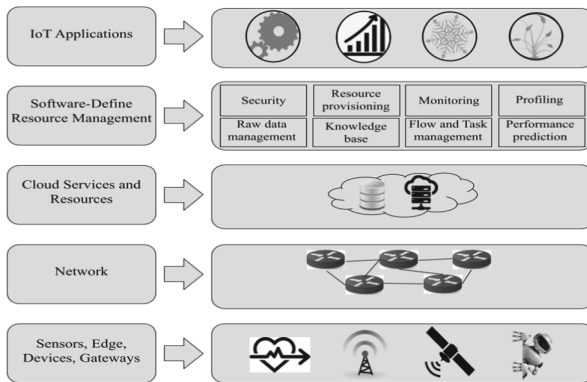
Most requests of smart devices pertain to its surroundings due to the high number of smart devices over the internet. For example, a user is allowed to identify individuals playing a similar sport nearby with the help of the aforementioned sports tracker application Endomondo. The request being processed in fog are reasonable compared to the cloud infrastructure due to the of the characteristic of the typical application request.

#### Reduction of Network Traffic

Since there are billions of devices present over the internet in today's date to send, receive, and generate data, it is neither sensible nor efficient to upload all the data to cloud. It is reasonable to put the computing capabilities near the location of devices, compared to sending the data over networks to central data centers. Therefore, a platform that filters and analyzes generated data closer to the edge is necessary.

## 2.2 Architecture of Fog Computing

The edge devices, end devices, and gateways are located at the bottommost layer. Apps that are installed to enhance functionality in end devices are located in this layer. Elements communicate among themselves, and with the cloud through the help of the network layer, suited above the bottommost layer. IoT tasks' processing and resource management are supported by the cloud services and resources situated in the cloud layer above the network layer. The resource management software lays above the cloud layer that offers quality of service to applications of fog computing and manages the entire infrastructure. The topmost layer delivers intelligent and innovative applications to end users by the applications in this layer which leverage fog computing. The architecture of fog computing is illustrated in Fig. 1 below.



**Fig. 1.** Architecture of Fog Computing

## 3 Homomorphic Encryption Scheme and Its Types

The concept of homomorphic encryption is obtained from abstract algebra. A combination of multiplication and addition Boolean circuits is used to create any kind of homomorphic encryption scheme. Such combinations are made of up AND and XOR gate.

Considering plaintext  $x_1, x_2, \dots, x_n$  and  $E$ , which is the scheme for encryption, is considered to be homomorphic in nature if it aligns with the additive property and/or multiplicative property. The different types of homomorphic encryption techniques are discussed in the subsections below.

### 3.1 Partially Homomorphic Encryption Scheme

On a ciphertext, either addition or multiplication can be performed with partially homomorphic encryption. Consider  $P$  as an encryption function, and two encrypted texts  $F(X)$  and  $F(Y)$ . Through the additive homomorphic approach, the function  $P$  can perform

some manipulations. Therefore, the resultant value is:  $F(X) + F(Y) = F(X + Y)$ . And similarly, the resultant value for the multiplicative homomorphic approach is:  $F(X) * F(Y) = F(X*Y)$ . Applications such as e-voting system is where partial homomorphic encryption schemes are extremely helpful since the votes can be counted without the decryption of original values. Therefore, secrecy and privacy are maintained. The multiplicative homomorphic approach is implemented by the RSA encryption, and Paillier encryption method implements the additive homomorphic approach.

### Paillier Encryption Scheme

In 1999, Pascal Paillier proposed the Paillier Encryption scheme. In this scheme, the additive homomorphism is described. A function  $F$  is used to encrypt two plain texts  $p_1$  and  $p_2$ .  $F(p_1 + p_2(n))$  is the resultant expression. The Paillier encryption algorithm has three functions: key generation, Encryption, Decryption [11].

*Key Generation:* In this function, two large prime numbers,  $p$  and  $q$ , and a security parameter  $g$  are selected. Here  $|p| = |q| = g$ . Next, the two keys, public and secret key, need to be evaluated. The public key,  $PbA = (Q, k)$ , where  $Q = pq$  and  $k \in Z_{N^2}^*$ . The secret key,  $\lambda$ , is evaluated from the lcm of  $(p - 1, q - 1)$ . Now, the public key, secret key, and the security parameter have been obtained.

*Encryption:* In order to generate the encrypted text from a given Message  $M$ , we use the formula:  $C = g^{Mr} \pmod{N^2}$ . Where  $r$  is a random number that belongs to  $Z_N^*$ . Now the encrypted text or the cipher text,  $C$ , has been obtained.

*Decryption:* The cipher text,  $C$ , which belongs to  $Z_{N^2}^*$ , is then decrypted to obtain the original message  $M$  using the formula:  $M = \frac{L(C^\lambda \pmod{N^2})}{L(g^\lambda \pmod{N^2})} \pmod{N}$ .

### RSA Encryption Scheme

In 1977, RSA algorithm was introduced [12]. The RSA algorithm implements the block cipher method. In cryptography, RSA algorithm is considered quite an efficient algorithm. Similar to Paillier encryption scheme, RSA also has three functions: key generation, encryption, and decryption.

*Key Generation:* Select  $p$  and  $q$ , which are two large prime numbers. Calculate  $n$ , where  $n = p * q$ . Then calculate,  $\phi(n) = \phi(p) * \phi(q)$ , therefore  $\phi(n) = (p - 1) * (q - 1)$ . Further, choose  $e$ , where  $1 < e < \phi(n)$ , and is co-prime to  $\phi(n)$ . Hence, the public key is  $(e, n)$ , and the private key is  $(d, n)$ , where  $d = e^{-1} \pmod{\phi(n)}$ .

*Encryption:* The cipher text,  $C$ , is calculated by the formula:  $C = P^e \pmod{n}$ , where  $P < n$ .

*Decryption:* The Plain Text,  $P$ , Can Be Obtained Through the Formula:  $P = C^d \pmod{n}$ .

## 3.2 Fully Homomorphic Encryption Scheme (FHE)

If an encryption technique supports both additive and multiplicative properties of homomorphism, then it is said to be fully homomorphic. In the year 2010, a lattice-based cryptosystem was proposed by Craig Gentry [13]. The cryptosystem was used to effectively

outsource the data. To modify in such a way that it only includes Boolean operations ( $\&$ ) and ( $\oplus$ ), the Gentry's scheme is implemented. For example, consider a Boolean expression  $X$ , expressed as  $X(\oplus)1$ . Similarly, the expression  $X \& Y$  is expressed as  $X(\oplus)1 \& Y(\oplus)1$ .

## 4 Various Homomorphic Encryption Schemes

This section reviews three homomorphic encryption schemes which are suitable to be implemented in Fog Computing: Brakerski-Gentry-Vaikuntanathan Model (BGV), Gorti's Improved Homomorphic Cryptosystem (GIHC), and ElGamal based Algebraic Homomorphic Encryption scheme (EAHE).

### 4.1 BGV

BGV provides ideal lattices fundamentals to form: Learning with Errors (LWE) and Ring-Learning with Errors (R-LWE) [14, 15]. The severity of decisional R-LWE is linked to data protection, along with processing of integer polynomials.

The BGV model follows includes various functions such as encryption, decryption, level shifting operations, and homomorphic operations. The sequence of these functions is set in a manner to define the process of BGV model. Firstly, the encryption is performed to encrypt the message with the help of the public key. Secondly, to obtain the original message, the cipher text is decrypted with the help of the private key. Thirdly, level shifting operations are performed to maintain the encrypted message dimensions, and reduce the noise. Lastly, homomorphic operations are performed in order to execute any addition and multiplication operations.

### 4.2 GIHC

In actuality, HME scheme of this type has a variety of applications. The primary presumption of GIHC is that the algorithm will change the current cipher text without disclosing the initial value. As the output information, the information that is encrypted will be returned, where the encryption method has a predefined setting.

### 4.3 EAHE

The EAHE is a framework of the NIST digital signature standard. IND CPA's enhanced security level is used in EAHE. Since the numbers generated at random are not the same, the receiver can't decipher the sender's encrypted message with EAHE, however, the sender is only able to decipher the encrypted text if it employs an identical key for both, encryption and decryption. The two users can separately encrypt their messages. It has been shown that EAHE is secured and resistant to plaintext attacks.

---

**Algorithm 1** EAHE Encryption

---

**Input:**  $p, q, M$ , where  $p$  and  $q$  are two large prime numbers, and  $M$  is the original message**Output:**  $Eg(M)$ , encrypted form of  $M$ 

1.  $N = p * q$
  2. Select an integer at random,  $x$ , and root value  $g$  of  $GF(p)$ , where  $g < p$  and  $x < p$
  3.  $y = (g * x) \bmod p$
  4.  $E1(M) = (M + r * p) \bmod N$
  5. Select random integer  $k$
  6.  $Eg(M) = (a, b) = ((g * k) \bmod p, (y * k) * E1(M) \bmod p)$
- 

---

**Algorithm 2** EAHE Decryption

---

**Input:**  $a, b$ , which are the  $x$  and  $y$  coordinate of the encrypted message on a coordinate plate**Output:**  $M$ , the original message

1.  $Dg(M) = b(ax) - 1 \pmod{p}$
- 

## 5 Role of Homomorphic Encryption Schemes in Fog Computing

Homomorphic encryption schemes are applicable in different sectors such as Healthcare, Banking, Government and Public, and E-commerce and online transactions [16, 17]. Various applications of Homomorphic encryption schemes in different sectors are discussed in the sub sections below and shown in Table 1.

### 5.1 Healthcare

While preserving patient privacy, homomorphic encryption can securely analyze and process sensitive electronic medical records. Without compromising patient confidentiality, homomorphic encryption secures data sharing among different organizations, as well as conduct collaborative research among healthcare institutions [18].

### 5.2 Banking

Homomorphic encryption provides for analyzing any financial requirements, or client loan application for predictive models based on the personal identity and behavior of customers. It ensures the privacy and confidentiality during the analyzing of data for a customer.

### 5.3 Government and Public

It's crucial to maintain sensitive citizen information when it comes to government and public sectors. Homomorphic encryption schemes can be applied in government sectors for secure data processing and analyzing of citizen information.



#### 5.4 E-Commerce and Online Transactions

An E-commerce platform consists of heavy user traffic, and to ensure the privacy and confidentiality of that user is crucial [19]. Therefore, homomorphic encryption is able to protect the customer information against unauthorized access and data breaches.

**Table 1.** Applications of Various Homomorphic Encryption Schemes

Encryption Scheme	Homomorphic Property	Applications
EAHE	Both (additive & multiplicative)	Mobile encryption, and e-voting system
GIHC	Both (additive & multiplicative)	Preserve privacy in message communication of MANET
BGV	Both (additive & multiplicative)	Preserving privacy in integer polynomials
RSA	Multiplicative	Security for banking transactions, and internet security
Paillier	Additive	Secure data aggregation in healthcare, E-polling system

## 6 Conclusion and Future Research Directions

The exponential rise of fog computing has brought more efficient and faster data processing. However, security of data has not experienced the same exponential rise. Therefore, data security in fog computing devices needs to be more enhanced. A detailed 5-layered architecture of fog computing has been illustrated with an overview of each layer that explains the working and functionality of the entire system. In order to evaluate the security parameters and identify areas that require protection in fog computing, we explored the various characteristics such as future scalability, low-latency, great computability with IoT devices, and network traffic reduction. This work identifies various homomorphic encryption schemes such as BGVM, GIHC, and EAHE that can be implemented to protect data and offer data security in fog computing devices. Further, various applications in different sectors such as healthcare, banking, government and public, and e-commerce have been explored to identify the significant role of homomorphic encryption schemes in fog computing.

Future research can be done towards analyzing and developing of more homomorphic encryption schemes that fit better and are more suitable for end-devices in fog computing. Other areas of improvement include, efficient key generation, faster encryption and decryption algorithms without the loss of security standards. Since, for most homomorphic encryption schemes the complexity and the implementation cost is very high, development of HME schemes with less complexity and cost can be another area of future research.



## References

1. Varshney, M., Bhushan, B., Haque, A.K.M.B.: Big Data Analytics and Data Mining for Healthcare Informatics (HCI). In: Kumar, R., Sharma, R., Pattnaik, P.K. (eds.) *Multimedia Technologies in the Internet of Things Environment*, Volume 3. *Studies in Big Data*, vol 108. Springer, Singapore (2022). [https://doi.org/10.1007/978-981-19-0924-5\\_11](https://doi.org/10.1007/978-981-19-0924-5_11)
2. Srivastava, S., Saxena, S., Buyya, R., Kumar, M., Shankar, A., Bhushan, B.: CGP: Cluster-based gossip protocol for dynamic resource environment in cloud. *Simul. Model. Pract. Theory* **108**, 102275 (2021). <https://doi.org/10.1016/j.simpat.2021.102275>
3. Diffie, W., Hellman, M.: New directions in cryptography. *IEEE Trans. Inf. Theory* **22**(6), 644–654 (1976)
4. Varshney, T., Sharma, N., Kaushik, I., Bhushan, B.: Authentication & Encryption Based Security Services in Blockchain Technology. In: 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) (2019). <https://doi.org/10.1109/iccis48478.2019.8974500>
5. Coron, J.S., Lepoint, T., Tibouchi, M.: Practical multilinear maps over the integers. In: *Annual Cryptology Conference*, pp. 476–493. Springer, Berlin, Heidelberg (2013, August)
6. Rivest, R.L., Adleman, L., Dertouzos, M.L.: On data banks and privacy homomorphisms. *Foundations of Secure Computation* **4**(11), 169–180 (1978)
7. ElGamal, T.: A public key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Trans. Inf. Theory* **31**(4), 469–472 (1985)
8. Paillier, P.: Public-key cryptosystems based on composite degree residuosity classes. *Eurocrypt* **99**, 223–238 (1999, May)
9. Bajaj, K., Sharma, B., Singh, R.: Comparative Analysis of Simulators for IoT Applications in Fog/Cloud Computing. In: 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), pp. 983–988. Erode, India (2022). <https://doi.org/10.1109/ICSCDS53736.2022.9760897>
10. Liu, C., Xiang, F., Wang, P., Sun, Z.: A review of issues and challenges in fog computing environment. In: 2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCCom/CyberSciTech), pp. 232–237. Fukuoka, Japan (2019). <https://doi.org/10.1109/DASC/PiCom/CBDCCom/CyberSciTech.2019.00050>
11. Zhang, Y., et al.: Privacy-preserving data aggregation against false data injection attacks in fog computing. *Sensors* **18**(8), 2659 (2018)
12. Stallings, W.: *Cryptography and network security: principles and practice*, international edition: principles and practice. Pearson Higher Ed (2014)
13. Gentry, C.: Fully homomorphic encryption using ideal lattices. In: *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pp. 169–178 (2009, May)
14. Lyubashevsky, V., Peikert, C., Regev, O.: On ideal lattices and learning with errors over rings. *J. ACM (JACM)* **60**(6), 1–35 (2013)
15. Fau, S., Sirdey, R., Fontaine, C., Aguilar-Melchor, C., Gogniat, G.: Towards practical program execution over fully homomorphic encryption schemes. In: 2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, pp. 284–290. IEEE (2013, October)
16. Abdalla, S., Hassan, M.S., Landolsi, T., Abu-Rukba, R.: On the performance of fog-cloud computing for real-time surveillance applications. In: 2020 International Conference on Communications, Signal Processing, and their Applications (ICCSPA), pp. 1–4. Sharjah, United Arab Emirates (2021). <https://doi.org/10.1109/ICCSPA49915.2021.9385745>

17. Kumar, M., Dubey, K., Pandey, R.: Evolution of emerging computing paradigm cloud to fog: applications, limitations and research challenges. In: 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 257–261. Noida, India (2021). <https://doi.org/10.1109/Confluence51648.2021.9377050>
18. Goyal, S., Sharma, N., Bhushan, B., Shankar, A., Sagayam, M.: Iot enabled technology in secured healthcare: Applications, challenges and future directions. *Cognitive Internet of Medical Things for Smart Healthcare*, 25–48 (2020). [https://doi.org/10.1007/978-3-030-55833-8\\_2](https://doi.org/10.1007/978-3-030-55833-8_2)
19. Bhushan, B., Kadam, K., Parashar, R., Kumar, S., Thakur, A.K.: Leveraging blockchain technology in sustainable supply chain management and logistics. In: Muthu, S.S. (ed.) *Blockchain Technologies for Sustainability. Environmental Footprints and Eco-design of Products and Processes*. Springer, Singapore (2022). [https://doi.org/10.1007/978-981-16-6301-7\\_9](https://doi.org/10.1007/978-981-16-6301-7_9)



# Automated and Optimised Machine Learning Algorithms for Healthcare Informatics

Aayush Juyal<sup>1</sup> , Bharat Bhushan<sup>1</sup> , Alaa Ali Hameed<sup>2</sup>, Akhtar Jamil<sup>3</sup>,  
and Shrayiash Pandey<sup>1</sup>

<sup>1</sup> Department of Computer Science and Engineering, School of Engineering and Technology, Sharda University, Greater Noida, Uttar Pradesh 201310, India

aayushjuyal12@gmail.com, bharat\_bhushan1989@yahoo.com,  
shrayiash.pandey@gmail.com

<sup>2</sup> Department of Computer Engineering, Istinye University, Istanbul, Turkey  
alaa.hameed@istinye.edu.tr

<sup>3</sup> Department of Computer Science, National University of Computer and Engineering Sciences, Islamabad, Pakistan  
akhtar.jamil@nu.edu.pk

**Abstract.** Healthcare is a rapidly expanding field with a substantial amount of heterogeneous data driving innumerable health-related tasks. Various healthcare service providers still rely on manual procedures, which can be time-consuming and require significant effort. To automate such manual operations, recent technological advances have emerged in the domain of Machine Learning (ML). ML falls under the subject of Artificial Intelligence (AI), and it gets combined with ‘big data’ to draw meaningful insights. The integration of ML in the healthcare sector has optimized decision-making and predictive analysis. This paper discusses the various application areas of ML in healthcare. Additionally, several ML algorithms used by other researchers in healthcare-related experiments are summarized. A brief review is provided regarding the experiments. This paper delineates the challenges associated with using ML in healthcare. Finally, the paper offers insights into future research directions.

**Keywords:** Healthcare · Machine Learning · Artificial Intelligence · Optimization · Automation · Security

## 1 Introduction

It is no surprise that healthcare is vital for every individual. It is a quickly growing sector, attracting scientists and medical professionals to contribute to the healthcare system. Healthcare facilities such as telemedicine, medical tools, health insurance, etc., can be enhanced to a great extent through progressive technologies. Thus, the progression of sophisticated technologies has generated copious data, which traditional methods fail to deal with. Traditional methods could involve medical professionals manually analyzing reports and medical images which could consume a fair amount of time. Hence, ML methods have proven to be a cutting-edge technology helping to improve healthcare

[1]. As numerous fields produce unimaginable amounts of data, ML methods can be considered the remedy of such a problem by processing complex and unstructured data. They extract meaningful insights from raw data that would be strenuous for a human to analyze manually, thereby harnessing the power of data and automation. ML is a concept introduced in the 1950s, and it has recently gained attention in fields including agriculture [2], finance [3], education [4], and healthcare [5]. Multiple fields use a suite of algorithms and statistical techniques to perform predictive analysis and optimize decision-making. ML's ability to approach a problem with different approaches suggests that it is flexible in handling all sorts of raw data, thereby making it applicable in numerous domains. It has been reported that more than 86% of healthcare service providers have adopted ML methods. The majority of its applications involve it being utilized for disease prediction, patient monitoring, and disease detection. For example, Monika et al. [6] detected and classified various skin cancers via ML and image-processing tools. Diwakar et al. [7] reviewed a list of ML algorithms and image fusion to help medical professionals predict heart disease. Tandel et al. [8] used both ML and deep learning (DL) methods to optimize brain tumor classification. Thus, it is expected that medical professionals like radiologists and physicians will be assisted by intelligent systems helping to revolutionize medical research and practice. With several experiments conducted by researchers, this paper sheds light on a diverse range of ML techniques and their appeal in healthcare. The following are the major contributions of this work.

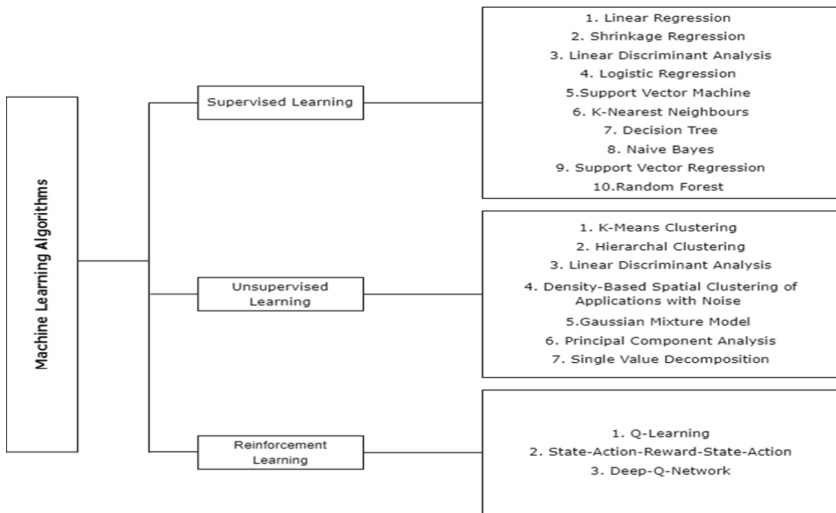
- This paper presents a brief overview of different ML algorithms, which are categorized into three types of learning, namely, supervised, unsupervised, and reinforcement.
- This paper outlines the issues in healthcare that are overcome by ML.
- This paper describes the four application areas of ML in healthcare, namely, prognosis, diagnosis, treatment, and clinical workflow.
- This paper reviews the work of other researchers that employed ML to solve healthcare issues.
- This paper discusses the challenges associated with using ML in healthcare.

The rest of the paper is structured as follows. Section 2 summarizes different ML algorithms. Section 3 discusses the applications of ML in healthcare. Section 4 analyses recent advances of ML in healthcare. Section 5 outlines the challenges that arise when ML is integrated with practical healthcare. Section 6 concludes this manuscript, followed by the future scope of work.

## 2 Background of Machine Learning Algorithms

ML is a vital discipline in the field of AI. Arthur Samuel, the AI pioneer, defined ML as the machine's ability to imitate intelligent human behavior without explicitly being programmed [9]. ML models are trained on a dataset to map the issue at hand to the principle of inference through statistical and logical techniques. While humans can perform the same tasks, it should be taken into consideration that humans are more prone to making mistakes because of their restricted short-term memory. With an excessive amount of data being generated in various fields and the growing ability of machines to process such data to draw up ML algorithms, humans can be aided by machines to

draw conclusions. ML algorithms can be classified into supervised, unsupervised, and reinforcement learning. Figure 1 delineates various ML algorithms.



**Fig. 1.** Categorization of machine learning algorithms

## 2.1 Supervised Learning

Algorithms working on labeled training data that learn a mapping function to determine the relationship between the input and output are referred to as supervised learning algorithms. The given algorithm is tested by providing it with unseen data to check whether it can recall the patterns it learned while training to make a prediction. The two sub-domains of supervised learning are regression and classification. Regression analysis focuses on fitting a function, from a family of functions to the sampled data, under a given loss function. Classification assigns a pre-determined label to an observation based on the features of observation. Table 1 outlines a plethora of supervised learning algorithms.

## 2.2 Unsupervised Learning

While supervised learning works on labeled data, it may not always be possible to get hold of data that is labeled, and assigning a label to each data instance can be tedious and time-consuming. Thus, unsupervised learning algorithms were introduced that operate on unlabelled data where patterns are identified by determining the similarities between the data points [15]. Clustering is a prime example, where, given a dataset  $X = \{x_1, x_2, x_3, \dots, x_n\}$ , can partition the dataset into  $K$  clusters,  $C = \{C_1, C_2, C_3, \dots, C_k\}$ . Dimensionality reduction algorithms are also a subset of unsupervised learning as it formulates a mathematical function  $f(X)$  to transform the data in a higher-dimensional space to a lower-dimensional representation. Table 2 summarizes the unsupervised learning algorithms.

**Table 1.** Summary of supervised learning algorithms

Supervised Learning Algorithm	Algorithm Description
Linear Regression	<ul style="list-style-type: none"> <li>• Finds the most optimal hyperplane to fit all the data points while keeping the sum of mean-squared error (MSE) minimum</li> <li>• Given N data points <math>(x_i, y_i)</math> where <math>x_i, y_i \in \mathbb{R} \forall i \in \{0, 1, \dots, P-1\}</math>, fit a linear function <math>\hat{y} = \beta_0 + \beta_1 x</math> (1) where <math>\hat{y}</math> is the output value, <math>\beta_0</math> is the y-intercept and <math>\beta_1</math> is the slope</li> </ul>
Shrinkage Regression	<ul style="list-style-type: none"> <li>• A statistical method to prevent overfitting in situations where a large number of predictors are present [10]</li> <li>• Adds a penalty term to the objective function, resulting in the coefficients' value approaching towards zero</li> </ul>
Linear Discriminant Analysis (LDA)	<ul style="list-style-type: none"> <li>• Discovers a linear combination of features that maximizes the separation between the classes</li> <li>• A hyperplane is drawn to separate the classes such that it maximizes the inter-class distance and minimizes the intra-class distance</li> </ul>
Logistic Regression (LR)	<ul style="list-style-type: none"> <li>• Used in binary classification to model the probability of a discrete result provided an input variable</li> <li>• Maps the probability of an event using sigmoid function [11]</li> <li>• <math>p(x) = \frac{1}{1+e^{-x}}</math> (2) where x is the input feature</li> </ul>
Support Vector Machine (SVM)	<ul style="list-style-type: none"> <li>• Formulates an optimal hyperplane in an N-dimensional space such that it maximizes the margin [12]</li> <li>• Hard margin and soft margin are the two techniques to determine the optimal hyperplane</li> </ul>
K-Nearest Neighbours (KNN)	<ul style="list-style-type: none"> <li>• A non-parametric algorithm that classifies the data points based on attributes</li> <li>• A new point <math>x_t</math>, where <math>X = \{x_1, x_2, x_3, \dots, x_n\}</math> is categorized by assigning it to an arbitrary class 'i' where majority of the neighbouring points lie, with respect to <math>x_t</math></li> </ul>
Decision Tree (DT)	<ul style="list-style-type: none"> <li>• A categorical value is predicted by recursively partitioning the decision space, resulting in the decision space becoming smaller, and then labels are assigned to this space</li> <li>• Models decisions and their result by following a tree-like structure from root node to leaf node</li> </ul>
Naïve Bayes (NB)	<ul style="list-style-type: none"> <li>• Classification algorithm that revolves around Bayes' theorem, with the assumption that all features are independent of one another</li> <li>• Objective is to choose a class Y where <math>Y \in \{0, 1\}</math>, such that posterior probability is maximized [13]</li> </ul>

*(continued)*

**Table 1.** (continued)

Supervised Learning Algorithm	Algorithm Description
Support Vector Regression (SVR)	<ul style="list-style-type: none"> <li>• Offers the flexibility of describing the amount of error acceptable by the model and thus finding a hyperplane that fits the data</li> <li>• Defines an objective function that minimizes the l2 norm of a coefficient vector</li> </ul>
Random Forest (RF)	<ul style="list-style-type: none"> <li>• An ensemble classification algorithm that is formulated by a collection of DTs</li> <li>• Results from all the DTs are aggregated to provide a prediction [14]</li> </ul>

**Table 2.** Outline of unsupervised learning algorithms

Unsupervised Learning Algorithm	Algorithm Description
K-Means Clustering	<ul style="list-style-type: none"> <li>• Iteratively assigns the data points to K centroids via a distance metric and updates the centroids</li> </ul>
Hierarchical Clustering	<ul style="list-style-type: none"> <li>• Iteratively merges or splits the clusters to build a tree-like structure referred to as dendrograms</li> <li>• Two types namely, agglomerative and divisive</li> </ul>
Density-Based Spatial Clustering of Applications with Noise (DBSCAN)	<ul style="list-style-type: none"> <li>• Groups data points in such a manner that separates the high density of points from the low density of points</li> <li>• Intuition is that for an arbitrary data point <math>x_i</math>, a minimum number of data points, 'minPts', must exist within the radius <math>\epsilon</math> of <math>x_i</math> for it to be considered as a cluster</li> </ul>
Gaussian Mixture Model (GMM)	<ul style="list-style-type: none"> <li>• Models a dataset with the assumption that clusters conform to Gaussian distribution</li> <li>• Discovers K Gaussian components to model the underlying distribution of the dataset</li> </ul>
Principal Component Analysis (PCA)	<ul style="list-style-type: none"> <li>• Dimensionality reduction algorithm that discovers various orthogonal axes (principal components) that captivate the largest variance</li> <li>• Mathematical techniques to compute principal components include covariance computation and eigenvalues and eigenvectors</li> </ul>
Single Value Decomposition (SVD)	<ul style="list-style-type: none"> <li>• Matrix factorization technique aiming to decompose a feature matrix into singular values</li> <li>• Given a matrix A, SVD decomposes it into <math>A = UWV^T</math> (3)</li> </ul> <p>where <math>U = AA^T</math>, W is the diagonal matrix and <math>V^T = A^T A</math>.</p>



### 2.3 Reinforcement Learning

A framework proposed in the early 1960s studied the way machines gained knowledge when a cumulative reward is awarded based on decisions taken by the machine. It begins with the machine taking some action in an environment followed by suitable feedback. The subsequent actions are modified according to the feedback and this process continues until the machine produces desired outcomes. This framework allows the machines to autonomously learn through the trial-and-error method. Table 3 overviews a couple of reinforcement learning algorithms.

**Table 3.** Overview of reinforcement learning algorithms

Reinforcement Learning Algorithm	Algorithm Description
Q – Learning	<ul style="list-style-type: none"> <li>• Computes Q-value, representing the cumulative reward the machine earns after performing an action</li> <li>• For a given current state <math>s</math> and action <math>a</math>, Q-value is given by the modified Bellman equation</li> </ul> $Q^*(s, a) = E_{s'} \left[ r + \lambda \max_{a'} Q^*(s', a')   s, a \right] \quad (4)$ <p>where <math>E_{s'}</math> is the expectation, <math>\lambda</math> is the discount factor, <math>Q^*</math> represents the optimal Q-value, <math>r</math> refers to the reward, <math>s'</math> denotes the next state and <math>a'</math> denotes the next action the machine can take</p>
State-Action-Reward-State-Action (SARSA)	<ul style="list-style-type: none"> <li>• An on-policy algorithm insinuating that Q-value is learned depending on current policy action instead of greedy policy</li> </ul>
Deep-Q-Network (DQN)	<ul style="list-style-type: none"> <li>• Combines Q-Learning with deep neural networks to deal with high dimensional spaces</li> </ul>

## 3 Applications of Machine Learning in Healthcare

The healthcare sector is responsible for generating an excessive amount of heterogeneous information, thus making it strenuous for conventional methods to process it. ML methods are adopted to gain insight and perform some action with this data i.e., predictive analysis, and detection. Social media, wearable devices, genomic data, etc. are a few examples of heterogenous sources of data that play a role in healthcare data. Figure 2 illustrates the four application areas of ML in healthcare with Table 4 summarizing those application areas.

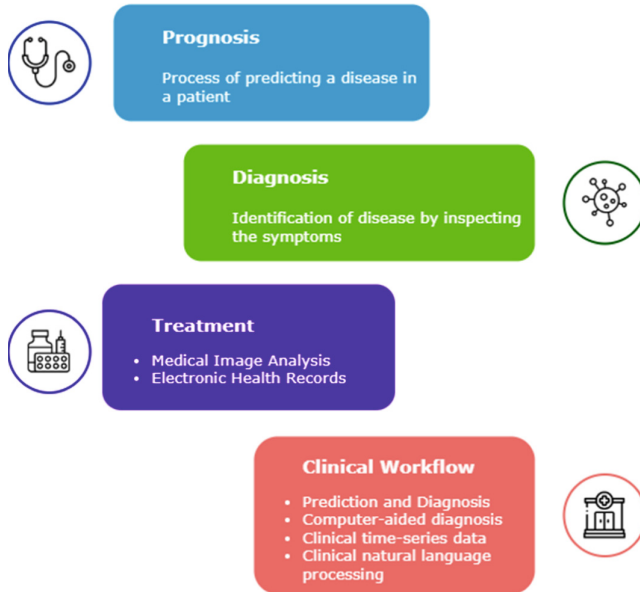


Fig. 2. Application areas of ML in healthcare

### 3.1 Prognosis

Prognosis refers to the procedure of predicting a potential disease that may arise in a patient. It estimates the progression of the disease in terms of whether it will aggravate, improve or remain as it is. The patient's symptoms are mapped to the specific disease helping in deciding which treatment plan should be employed. ML models are able to predict and classify diseases by working with data such as medical images, phenotypic samples, genotypic samples, and proteomic data, etc.

### 3.2 Diagnosis

Diseases can be identified by analysing and evaluating the patient's symptoms, medical past, and diagnostic test samples. Electronic health records (EHRs) are generated on a daily basis in hospitals consisting of data that represents the medical history of patients. ML models can extract relevant attributes from EHRs to ease the diagnosis process. Subsequently, medical image analysis leverage ML techniques by working with various imaging modalities namely, computed topography (CT), medical image resonance (MRI), ultrasound, etc. The predominant use of ML is to extract key characteristics in the image acquired which can be mapped to an expected disease [29].

### 3.3 Treatment

Personalized treatment recommendations can be suggested by healthcare professionals based on the disease the patient is suffering from. Healthcare professionals are assisted

through ML methods where they analyze large datasets describing the patient and the symptoms to provide valuable insight and help nudge the healthcare professionals in suggesting a treatment plan. As discussed in the above sub-section, medical images are widely used to help diagnose a potential disease. They get interpreted by radiologists and physicians which involves them writing a detailed report about each body organ. This can be arduous and time-consuming which results in introducing Natural language processing (NLP) and ML methods to tackle this issue [16, 17]. Real-time monitoring is another critical aspect of treatment. Wearable devices and Internet of Things (IoT) sensors are able to collect health-centric data which is transmitted over the cloud to be used by ML methods [18, 28].

### 3.4 Clinical Workflows

ML has significantly impacted and optimized various aspects of clinical workflows in different ways. Predictive analysis and disease diagnosing are one of the primary applications of ML in the healthcare industry. Computer-aided diagnosis has aided radiologists in automating the interpretation of medical images through ML and computer vision techniques. Another task of clinical workflow involves the analysis of time-series data that covers measurements such as vital signs, behavioral data, laboratory results, etc. Clinical text found in notes written by clinicians is vital for communicating the state of a patient. Finally, NLP techniques have been progressing to extract the key details from the unstructured representation of notes [30, 31].

**Table 4.** Summary of ML applications across healthcare

Problem Domain	ML Application
Prognosis	Predicts diseases by working with medical-generated data i.e., medical images
Diagnosis	Detects diseases based on EHRs and medical images
Treatment	Helps medical professionals devise a treatment plan based on the disease the patient is suffering from
Clinical Workflow	Disease detection, time-series analysis, clinical text analysis, disease diagnosis and prediction

## 4 Recent Advances

This section reviews the work of other researchers in terms of the four application areas as mentioned in Sect. 3. Table 5 summarizes the appeal of ML in the healthcare ecosystem.

Vaka et al. [19] presented a novel approach for breast cancer (BC) detection by implementing a deep neural network with support value (DNNS). Alickovic et al. [20] automated the detection of Alzheimer by using a histogram to compute feature vectors that acted as input to the RF classifier.

Assaf et al. [21] utilized three ML models namely, Artificial Neural Networks (ANN), RF, and classification and regression tree (CART) to predict the outcome of patients suffering from non-critical COVID-19. Nemesure et al. [22] evaluated the significance of data collected from EHRs on generalized anxiety disorder (GAD) and major depressive disorder (MDD) through several state-of-the-art ML models. Alzu'bi et al. [23] used a combination of NLP and ML techniques to extract relevant attributes of BC from EHRs to build a predictive model that estimates the probability of recurrence of BC.

Taalat [24] presented an optimized reinforcement learning (ORL) approach to optimize hyperparameters using particle swarm optimization (PSO) and subsequently use that algorithm for resource allocation algorithm (RAA) and load balancing in a Fog environment. Mohan et al. [25] predicted the daily COVID-19 cases during the third wave in India through the support of time-series forecasting models and NLP, which was primarily applicable in sentiment analysis. The different models employed include autoregressive integrated moving average (ARIMA), Susceptable-Exposed-Infectious-Recovered (SEIR), Polynomial Regression (PR), and Prophet.

Hennebelle et al. [26] proposed a smart healthcare framework that integrates ML for the prediction of type-2 diabetes in an IoT-based cloud computing environment. Yuan et al. [27] employed a deep-reinforcement-learning-based task-oriented homogenized automatic preprocessing model (DRL-HAPre) to obtain an accurate and rapid medical image analysis approach.

**Table 5.** Recent advances of ML in healthcare

Reference	Year	Major Contribution	ML Algorithm(s)
Vaka et al. [19]	2020	Used the support value in a DNN for BC prediction	DNNS
Alickovic et al. [20]	2020	Presented a two-step framework for the quick and easy diagnosis of Alzheimer disease	RF
Assaf et al. [21]	2021	Determined the vulnerability for non-critical COVID-19 patients using clinical parameters	ANN, RF, CART
Nemesure et al. [22]	2021	Predicted MDD and GAD with the potential of identifying unknown mental disorders	XGBoost, RF, SVM, KNN, NN
Alzu'bi et al. [23]	2021	Predicted the chance of BC recurring	NB, SVM, KNN, Multi-Layer Perceptron (MLP), LR, Bagging

(continued)

**Table 5.** (continued)

Reference	Year	Major Contribution	ML Algorithm(s)
Taalat [24]	2022	Optimized RL hyperparameters for resource allocation	ORL, RAA
Mohan et al. [25]	2022	Leveraged a time-series model to predict the COVID-19 cases	ARIMA, SEIR, Prophet, PR
Hennebelle et al. [26]	2023	Used IoT devices in a cloud computing environment to predict type-2 diabetes	RF, LR
Yuan et al. [27]	2023	Solved the issue of medical image analysis solutions being too slow and inaccurate	DRL-HAPre

## 5 Challenges of Machine Learning in Healthcare

Though ML holds great potential for healthcare-based tasks, there are several drawbacks to using ML. Table 6 outlines the challenges associated with using ML in practical healthcare.

**Table 6.** Delineating challenges of ML in healthcare

Challenge	Challenge Description
Safety	<ul style="list-style-type: none"> <li>• Use of ML algorithms through external systems must be non-harmful to the user</li> <li>• Flawed predictions can result in misdiagnosis further leading to incorrect treatment plans thus compromising patient safety</li> </ul>
Privacy	<ul style="list-style-type: none"> <li>• ML models make use of confidential user data i.e., age, sex, date of birth etc</li> <li>• Potential threats could involve revelling such information or the unethical use of it</li> </ul>
Ethical	<ul style="list-style-type: none"> <li>• Ethical principles must be taken into consideration before applying ML models</li> <li>• A right balance between innovation and following ethical norms must be established</li> </ul>
Regulatory and Policy	<ul style="list-style-type: none"> <li>• Integration of the ML model must be in compliance with the rules and regulations prescribed by the government</li> <li>• Adhering to such rules and regulations can be tedious and time-consuming</li> </ul>

(continued)

**Table 6.** (continued)

Challenge	Challenge Description
Data	<ul style="list-style-type: none"> <li>• Availability of high-quality and diverse data resembling clinical environment is rare</li> <li>• Raw data collected in practice tends to be subjective, biased, and redundant, therefore ML working on such data will produce undesirable outcomes</li> </ul>
Interpretability	<ul style="list-style-type: none"> <li>• Complex ML models can be difficult to understand</li> <li>• It is necessary to understand how ML models arrived at that specific prediction or conclusion for the safety of the patients</li> </ul>

## 6 Conclusion and Future Research Direction

The healthcare sector is responsible for producing a considerable sum of data, making it increasingly difficult for conventional methods to process and analyze that data. Since most of these conventional methods are manual, time and efficiency have been compromised. The introduction of ML techniques has witnessed the transformation of manual conventional methods to more automated, time-conserving practices. ML has provided a range of intelligent algorithms for tasks like disease prediction, disease diagnosis, medical image analysis, etc. This manuscript briefs about several ML algorithms that are categorized into supervised, unsupervised, and reinforcement learning. The application of ML in sub-domains of the healthcare ecosystem has also been presented. Furthermore, the work of other researchers is reviewed, highlighting experiments aimed at revolutionizing prediction and decision-making in the healthcare sector. As a scope of future work, transfer learning models have been gaining significant attention from researchers where pre-trained models have the potential use across multiple health institutions while preserving the privacy and security of the patient.

## References

1. Kumari, J., Kumar, E., Kumar, D.: A structured analysis to study the role of machine learning and deep learning in the healthcare sector with big data analytics. *Arch. Computat. Methods Eng.* **30**, 3673–3701 (2023). <https://doi.org/10.1007/s11831-023-09915-y>
2. Kansal, N., Bhushan, B., Sharma, S.: Architecture, Security Vulnerabilities, and the Proposed Countermeasures in Agriculture-Internet-of-Things (AIoT) Systems. In: Pattnaik, P.K., Kumar, R., Pal, S. (eds.) *Internet of Things and Analytics for Agriculture*, Volume 3. *Studies in Big Data*, vol 99. Springer, Singapore (2022). [https://doi.org/10.1007/978-981-16-6210-2\\_16](https://doi.org/10.1007/978-981-16-6210-2_16)
3. Goodell, J.W., Kumar, S., Lim, W.M., Pattnaik, D.: Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. *J. Behav. Exp. Financ.* **32**, 100577 (2021). <https://doi.org/10.1016/j.jbef.2021.100577>
4. Qazdar, A., Er-Raha, B., Cherkaoui, C., et al.: A machine learning algorithm framework for predicting students performance: a case study of baccalaureate students in Morocco. *Educ. Inf. Technol.* **24**, 3577–3589 (2019). <https://doi.org/10.1007/s10639-019-09946-8>

5. Goyal, S., Sharma, N., Bhushan, B., Shankar, A., Sagayam, M.: Iot enabled technology in secured healthcare: Applications, challenges and future directions. *Cognitive Internet of Medical Things for Smart Healthcare*, 25-48 (2020). [https://doi.org/10.1007/978-3-030-55833-8\\_2](https://doi.org/10.1007/978-3-030-55833-8_2)
6. Monika, M.K., Arun-vignesh, N., Usha Kumari, C., Kumar, M., Lydia, E.L.: Skin cancer detection and classification using machine learning. *Materials Today: Proceedings* **33**, 4266–4270 (2020). <https://doi.org/10.1016/j.matpr.2020.07.366>
7. Diwakar, M., et al.: Latest trends on heart disease prediction using machine learning and image fusion. *Materials Today: Proceedings* **37**, 3213–3218 (2021). <https://doi.org/10.1016/j.matpr.2020.09.078>
8. Tandel, G.S., Tiwari, A., Kakde, O.: Performance optimisation of deep learning models using majority voting algorithm for brain tumour classification. *Comput. Biol. Med.* **135**, 104564 (2021). <https://doi.org/10.1016/j.compbiomed.2021.104564>
9. Bhowmik, T., Bhadwaj, A., Kumar, A., Bhushan, B.: Machine learning and deep learning models for privacy management and data analysis in smart cities. In: Balas, V.E., Solanki, V.K., Kumar, R. (eds.) *Recent Advances in Internet of Things and Machine Learning*. Intelligent Systems Reference Library, vol 215. Springer, Cham (2022). [https://doi.org/10.1007/978-3-030-90119-6\\_13](https://doi.org/10.1007/978-3-030-90119-6_13)
10. Van Erp, S., Oberski, D.L., Mulder, J.: Shrinkage priors for Bayesian penalized regression. *J. Math. Psychol.* **89**, 31–50 (2019). <https://doi.org/10.1016/j.jmp.2018.12.004>
11. Deng, S., Wei, M., Xu, M., et al.: Prediction of the rate of penetration using logistic regression algorithm of machine learning model. *Arab. J. Geosci.* **14**, 2230 (2021). <https://doi.org/10.1007/s12517-021-08452-x>
12. Kumar, P., Hati, A.S.: Review on machine learning algorithm based fault detection in induction motors. *Archives of Computational Methods in Engineering* **28**(3), 1929–1940 (2021). <https://doi.org/10.1007/s11831-020-09446-w>
13. Wickramasinghe, I., Kalutarage, H.: Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. *Soft. Comput.* **25**, 2277–2293 (2021). <https://doi.org/10.1007/s00500-020-05297-6>
14. Speiser, J.L., Miller, M.E., Tooze, J., Ip, E.: A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst. Appl.* **134**, 93–101 (2019). <https://doi.org/10.1016/j.eswa.2019.05.028>
15. Wang, J., Biljecki, F.: Unsupervised machine learning in urban studies: A systematic review of applications. *Cities* **129**, 103925 (2022). <https://doi.org/10.1016/j.cities.2022.103925>
16. Tyagi, N., Bhushan, B.: Demystifying the Role of Natural Language Processing (NLP) in Smart City Applications: Background, Motivation, Recent Advances, and Future Research Directions. *Wireless Pers. Commun.* **130**, 857–908 (2023). <https://doi.org/10.1007/s11277-023-10312-8>
17. Tyagi, N., Bhushan, B.: Natural Language Processing (NLP) Based Innovations for Smart Healthcare Applications in Healthcare 4.0. In: Ahad, M.A., Casalino, G., Bhushan, B. (eds.) *Enabling Technologies for Effective Planning and Management in Sustainable Smart Cities*. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-22922-0\\_5](https://doi.org/10.1007/978-3-031-22922-0_5)
18. Varshney, M., Bhushan, B., Haque, A.K.M.B.: Big Data Analytics and Data Mining for Healthcare Informatics (HCI). In: Kumar, R., Sharma, R., Pattnaik, P.K. (eds.) *Multimedia Technologies in the Internet of Things Environment*, Volume 3. *Studies in Big Data*, vol 108. Springer, Singapore (2022). [https://doi.org/10.1007/978-981-19-0924-5\\_11](https://doi.org/10.1007/978-981-19-0924-5_11)
19. Vaka, A.R., Soni, B., Sudheer Reddy, K.: Breast cancer detection by leveraging Machine Learning. *ICT Express* **6**(4), 320–324 (2020). <https://doi.org/10.1016/j.ict.2020.04.009>
20. Alickovic, E., Subasi, A.: for the Alzheimer’s Disease Neuroimaging Initiative. Automatic Detection of Alzheimer Disease Based on Histogram and Random Forest. In: Badnjevic, A.,

- Škrbić, R., Gurbeta Pokvić, L. (eds.) CMBEBIH 2019. CMBEBIH 2019. IFMBE Proceedings, vol 73. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-17971-7\\_14](https://doi.org/10.1007/978-3-030-17971-7_14)
21. Assaf, D., Gutman, Y., Neuman, Y., et al.: Utilization of machine-learning models to accurately predict the risk for critical COVID-19. *Intern. Emerg. Med.* **15**, 1435–1443 (2020). <https://doi.org/10.1007/s11739-020-02475-0>
  22. Nemesure, M.D., Heinz, M.V., Huang, R., et al.: Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. *Sci. Rep.* **11**, 1980 (2021). <https://doi.org/10.1038/s41598-021-81368-4>
  23. Alzu'bi, A., Najadat, H., Doulat, W., et al. : Predicting the recurrence of breast cancer using machine learning algorithms. *Multimed Tools Appl* **80**, 13787–13800 (2021). <https://doi.org/10.1007/s11042-020-10448-w>
  24. Talaat, F.M.: Effective deep Q-networks (EDQN) strategy for resource allocation based on optimized reinforcement learning algorithm. *Multimed Tools Appl* **81**, 39945–39961 (2022). <https://doi.org/10.1007/s11042-022-13000-0>
  25. Mohan, S., Solanki, A.K., Taluja, H.K., Singh, A.: Predicting the impact of the third wave of COVID-19 in India using hybrid statistical machine learning models: a time series forecasting and sentiment analysis approach. *Comput. Biol. Med.* **144**, 105354 (2022). <https://doi.org/10.1016/j.combiomed.2022.105354>
  26. El-Hasnony, I.M., Elzeki, O.M., Alshehri, A., Salem, H.: Multi-label active learning-based machine learning model for heart disease prediction. *Sensors* **22**(3), 1184 (2022). <https://doi.org/10.3390/s22031184>
  27. Yuan, D., Liu, Y., Xu, Z., Zhan, Y., Chen, J., Lukasiewicz, T.: Painless and accurate medical image analysis using deep reinforcement learning with task-oriented homogenized automatic pre-processing. *Comput. Biol. Med.* **153**, 106487 (2023). <https://doi.org/10.1016/j.combiomed.2022.106487>
  28. Zafar, A., Wajid, M.A.: A mathematical model to analyze the role of uncertain and indeterminate factors in the spread of pandemics like COVID-19 using neutrosophy: a case study of India, Vol. 38. *Infinite Study* (2020)
  29. Wajid, M.A., Zafar, A., Bhushan, B., Khanday, A.M.U.D., Wajid, M.S.: Artificial Intelligence (AI) and Internet of Things (IoT): Application in Detecting and Containing the Spread of COVID-19. In: *AI Models for Blockchain-Based Intelligent Networks in IoT Systems: Concepts, Methodologies, Tools, and Applications*, pp. 373–392. Springer International Publishing, Cham (2023)
  30. Wajid, M.A., Zafar, A., Wajid, M.S., Terashima-Marín, H.: Neutrosophic-CNN-based image and text fusion for multimodal classification. *Journal of Intelligent & Fuzzy Systems* (Preprint), 1–17 (2023)
  31. Gupta, P.K., et al.: COVID-WideNet—a capsule network for COVID-19 detection. *Appl. Soft Comput.* **122**, 108780 (2022)





# A Survey on Image-Based Cardiac Diagnosis Prediction Using Machine Learning and Deep Learning Techniques

Anindya Nag<sup>1</sup> , Biva Das<sup>2</sup> , Riya Sil<sup>3</sup> , Alaa Ali Hameed<sup>4</sup>, Bharat Bhushan<sup>5</sup> ,  
and Akhtar Jamil<sup>6</sup>  

<sup>1</sup> Computer Science and Engineering Discipline, Khulna University, Khulna 9208, Bangladesh

<sup>2</sup> Pharmacy Discipline, Khulna University, Khulna 9208, Bangladesh

<sup>3</sup> Department of Computer Science and Engineering, Adamas University, Kolkata 700126, India

<sup>4</sup> Department of Computer Engineering, Istinye University, Istanbul, Turkey

alaa.hameed@istinye.edu.tr

<sup>5</sup> Department of Computer Science and Engineering, Sharda University, Greater Noida, Uttar Pradesh 201310, India

<sup>6</sup> Department of Computer Science, National University of Computer and Engineering Sciences, Islamabad, Pakistan

akhtar.jamil@nu.edu.pk

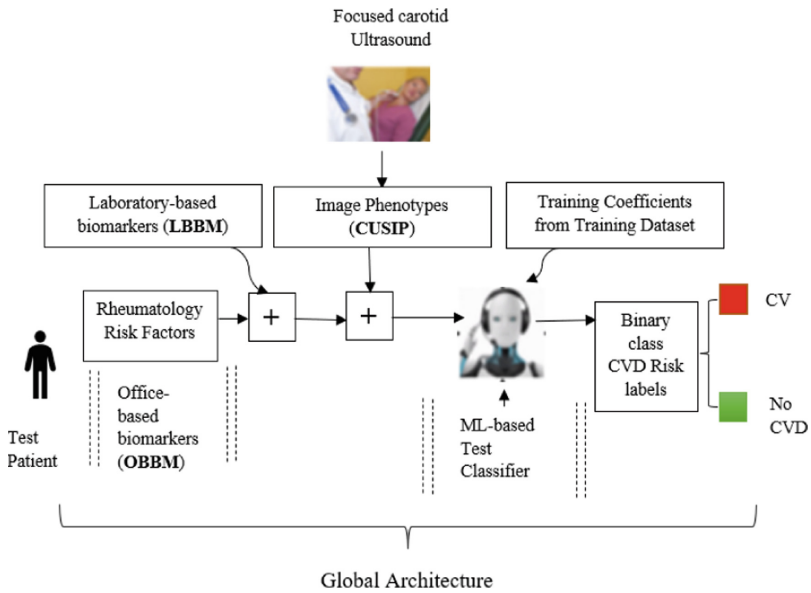
**Abstract.** Cardiac imaging is crucial in the diagnosis of cardiovascular disease. Cardiovascular disease is the umbrella term for the majority of heart ailments. The majority of the causes of mortality are associated with cardiovascular illness. The authors provide a technique for the diagnosis of cardiac disease. The main aim of this study is to determine the most effective technique for predicting cardiovascular disease, specifically focusing on the use of signs of heart disease and Electrocardiogram images. This will be achieved by leveraging the latest advancements in Deep Learning and Machine Learning methods. The authors conduct a comprehensive examination of various Machine Learning and Deep Learning Techniques. These techniques were evaluated in the context of predicting cardiovascular disease evaluating Image. The analysis shows that the Convolutional Neural Network methods are much more effective than the alternatives.

**Keywords:** Cardiovascular Disease · Deep Learning · Machine Learning · ECG images · Cardiac imaging · CNN

## 1 Introduction

The evaluation of massive data sets that could be used to forecast, prevent, manage, and cure diseases is a key function of the rapidly developing field of big data, which has played an increasingly important role in the administration of health care [1, 2]. Traditional Machine Learning (ML) approaches struggle when confronted with the 5 Vs (velocity, variety, volume, veracity, and value) of big data during massive data processing. ML approaches need to be updated so that they could be used for processing massive amounts of data [3]. To anticipate reliable patient data analysis, ML has been suggested for use in health analytics [4].

Cardiovascular disease (CVD) is a leading cause of death worldwide [5]. The most common forms of CVD include stroke, coronary heart disease, rheumatic heart disease, peripheral artery disease, and congenital heart defects. The World Health Organization (WHO) estimates that 17.9 million lives are lost annually owing to CVD and associated causes. Heart disease and cerebrovascular accidents account for almost 80% of all fatalities from CVD. Some of the possible risk factors that speed up heart-related issues include unhealthy diets, lack of physical exercise, alcohol misuse, and cigarette use. The outcome is the prevalence of intermediate-risk conditions including hypertension, diabetes, abnormal lipid profiles, excess body fat, and being overweight or obese [6]. However, early detection of patients at high risk of CVD and the provision of effective medications could avert unexpected and untimely deaths. The overall framework of the suggested online ML-based CVD risk classification system is shown in Fig. 1.



**Fig. 1.** General layout of the planned CVD risk stratification system based on ML [7]

The Electrocardiogram (ECG) is used to identify cardiovascular disease. Nevertheless, it requires time and effort to visually recognize long-term ECG abnormalities. Cardiovascular diseases are included in Table 1. Predicting CVDs requires analyzing ECG data together with other characteristics like age, sex, smoking status, alcohol use, blood pressure, cholesterol levels, etc. There are a wide variety of approaches that could be taken to analyze these variables. The study of CVD prediction has, in recent years, centered on the application of DL algorithms. Many academics and medical professionals have learned that ML-Based Heart Disease Diagnosis (MLBHDD) systems are low-cost and adaptable since the introduction of ML to the medical industry [8, 9]. Multiple investigations employing various CVD datasets, therefore, suggested MLBHDD [10]. Compared to more conventional learning algorithms, a Deep Learning (DL) based

Convolutional Neural Network (CNN) models perform well [11]. Figure 2 demonstrates this pattern since it depicts the dramatic rise in the number of studies using CNN for cardiac image segmentation over the last several years [12].

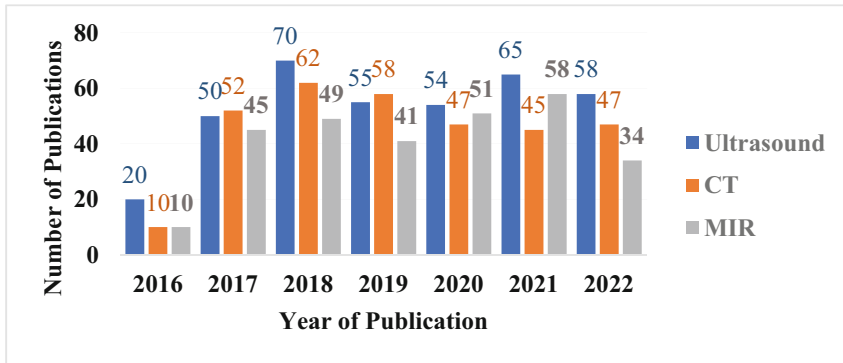


Fig. 2. Overview of DL-based cardiac image segmentation method published 2016–2022 [12]

Table 1 provides a concise categorization of CVD.

Table 1. Classification of CVD [13]

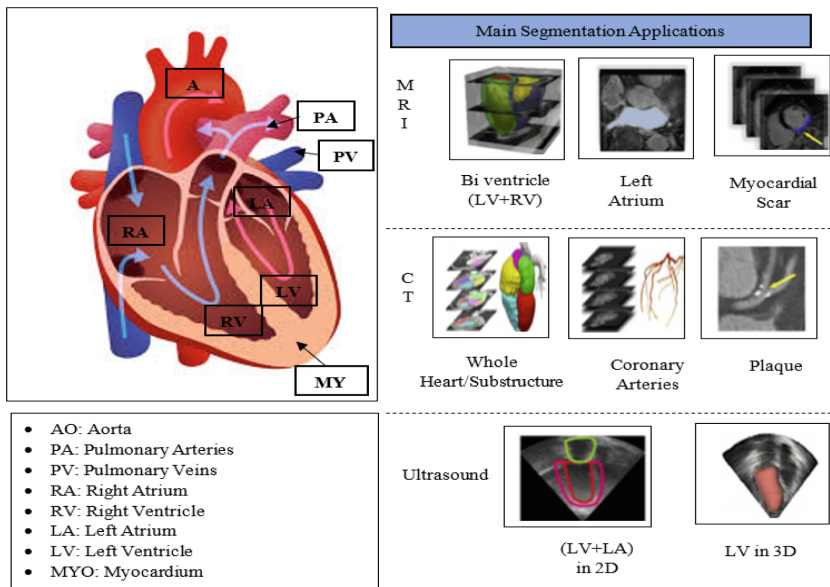
Name	Description	Possible risk factors
Coronary heart diseases	Diseases of the blood vessels	Conditions including high blood pressure, inactivity, diabetes, old age, blood clotting disease, cigarette use, poor nutrition, and so on are all contributors
Rheumatic heart disease	Broken heart valves and strained heart muscles	Disease caused by streptococcal bacteria; also known as rheumatic fever
Congenital heart disease	Inherited abnormalities in the heart’s structure	Consumption of alcoholic beverages, usage of some pharmaceuticals by the pregnant woman, etc
Aortic aneurysm and dissection	The phenomenon of aortic dilatation and rupture	Getting older, cardiac defects at birth, hypertension throughout time, syphilis, infectious/inflammatory diseases, and Marfan syndrome

(continued)

**Table 1.** (continued)

Name	Description	Possible risk factors
Peripheral arterial disease	Disorders affecting the blood vessels that supply the legs	CVD is attributed to a common set of factors
Deep Venous Thrombosis (DVT) and pulmonary embolism	Concentrations of blood in the legs, which could go to the abdominal organs and the lungs	Diabetes, cancer, deep vein thrombosis, recent delivery, the use of certain medications, hormone replacement therapy, and other risk factors
Arrhythmia	Atypical heart rhythm	Strong physical exertion might cause irregular heartbeats

Figure 3 represented a prediction-centric ensemble model by combining several ML techniques [14]. Figure 3 provides a high-level overview of frequent tasks associated with cardiac image segmentation, with examples for the three most prevalent modalities as Computed Tomography (CT), ultrasound and Magnetic Resonance Imaging (MRI).

**Fig. 3.** Description of Segmentation Tasks for Cardiac Images across Imaging Modalities [15].

In this study, the authors outline the most important ML approaches and the processes involved in improving and validating novel ML tools for image-based diagnosis. Additionally, the authors present a comprehensive literature analysis on the application of ML in the image-based diagnosis of CVD. The next four sections of the paper elaborate on these themes. Section 1 describes the fundamentals of ML; Sect. 2 is the

literature review. The literature evaluation is stated in Sect. 3, followed by a conclusion and potential future applications in Sect. 4.

### 1.1 Fundamental of Machine Learning (ML)

ML algorithms are classified as Supervised Learning (SL), Unsupervised Learning (UL), Semi-supervised Learning (SSL) and Reinforcement Learning (RL) [16]. The computer in SL is provided with a dataset that has been appropriately labelled. It includes parameters for input and output. Then, when a machine is given a new dataset, the SL algorithm analyses it in light of the labeled data and generates the appropriate output. In UL, there is no labeled dataset for the computer to use, and the algorithm is meant to point these things out on its own. The process involves categorizing information. Algorithms in RL are crafted in such a manner that the machine actively seeks the ideal solution, uses the reward and punishment principle to guide its actions, and eventually achieves the desired outcome. Logistic regression (LR), Support Vector Machine (SVM), Random Forests (RF), cluster analysis, Convolutional Neural Networks (CNN), and Artificial Neural Networks (ANN), are some of the most popular ML approaches utilized in cardiac imaging and diagnostics today [17, 18] (Fig. 4).

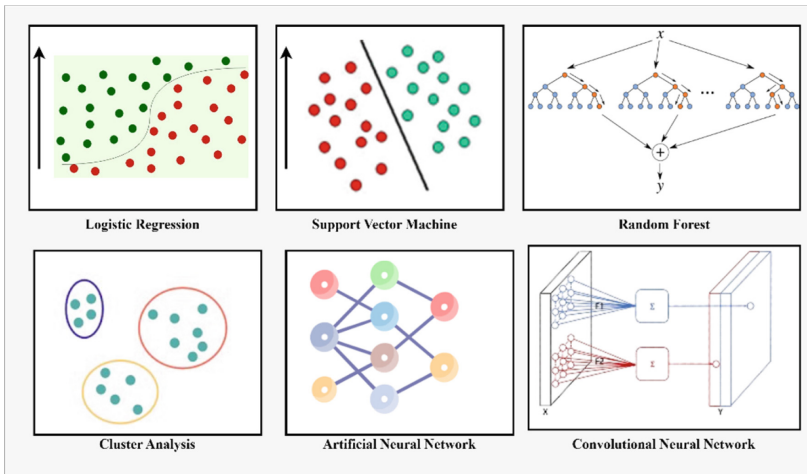


Fig. 4. The most popular ML methods currently being utilized in cardiac imaging [19].

**Logistic Regression (LR):** The ML method of logistic regression is called supervised learning. The sigmoid function (logarithmic) provides the foundation since it takes a real number and returns a value between 0 and 1. An N-sample training dataset is given to the algorithm during the training phase. Each instance has certain X properties and a Y label that specifies the way it should be categorized. As a result of the training process, the system generates a model that could be used to label data that was not involved in the training set [20, 21].

**Support Vector Machine (SVM):** The SVM is another popular cutting-edge ML method. Its primary function is to organize data. As an algorithm, SVM relies on the concept of margin calculation. Essentially, it establishes artificial barriers between economic classes. Classes are separated from the margin as much as possible to provide the lowest possible classification error in the process [20, 21].

**Random Forest (RF):** The RF method is a regression technique that utilizes ensemble learning by constructing several decision trees in the course of training. The output of a RF is the class chosen by the majority of trees, and this is useful for tasks requiring categorization. The average prediction of all the trees is all that is given back during regression tasks [22].

**Clustering Analysis:** Unsupervised cluster analysis identifies subgroups within the input data. Clustering algorithms group objects together depending on how far apart they are, thus the degree to which an item is a part of a set is measured by its distance from the set's center.

**Artificial Neural Network (ANN):** ANN is inspired by the way neurons operate in the human brain. Neurons are cellular-like structures found in the brain. Neuron theory is essential to understanding neural networks. Dendrites, nucleus, soma, and axon are their names. Artificial neural networks exhibit the same kinds of behavior. There are three levels at which it operates. Data is sent into the input layer (much like dendrites). The input is processed by the underlying layer, which is concealed (like soma and axon). Finally, the calculated output is sent by the output layer [23]. The three main categories of artificial neural networks are reinforcement, unsupervised, and supervised.

**Convolutional Neural Network (CNN):** CNNs are an extension of ANNs that use a process called convolutional product to adjust the value of a node in a given layer based on its spatial proximity to a node in the layer from above. These models were created specifically for use in image processing, where it is crucial that nodes' (pixels') spatial information be considered when making predictions. ANNs are given access to both the benefits and the downsides. These models vary significantly from their predecessors in that it takes images as input without doing any feature extraction.

Table 2 introduces the most popular ML approaches used in cardiac imaging and diagnostics and explores their advantages and disadvantages for a lay audience.

**Table 2.** Overview of ML methods [22, 23]

Techniques	Description	Advantages	Disadvantages
LR	Extending linear regression to produce a binary classification as the final output	<ul style="list-style-type: none"> <li>• 1. Easy to carry out</li> <li>• 2. Quick in terms of processing</li> </ul>	Does not make sense within the framework of the non-linearity model

(continued)

**Table 2.** (continued)

Techniques	Description	Advantages	Disadvantages
SVM	Determines the most appropriate divide between classes	<ol style="list-style-type: none"> <li>1. High accuracy</li> <li>2. Rapid Categorization of Recently Discovered Objects</li> </ol>	<ol style="list-style-type: none"> <li>1. The selection of the kernel might not be immediately apparent</li> <li>2. Computationally intensive</li> </ol>
RF	Creates a series of hierarchical decision queries that are executed on the input and output data	<ol style="list-style-type: none"> <li>1. Strong Performance</li> <li>2. Derive variable measures</li> </ol>	<ol style="list-style-type: none"> <li>1. Computation intensive</li> <li>2. Overfit problem</li> </ol>
ANN	Models perform complex classification tasks by passing input data through a series of nonlinear transformations in a network	<ol style="list-style-type: none"> <li>1. Handle noisy data</li> <li>2. Detect non-linear relation</li> </ol>	<ol style="list-style-type: none"> <li>1. Slow training time</li> <li>2. Hard to interpret</li> </ol>
CNN	ANNs that have been modified for image data processing and categorization	<ol style="list-style-type: none"> <li>1. daptable design based on the requirements of the application</li> <li>2. Capable of gaining knowledge of the optical characteristics straight from the images</li> </ol>	Same limitations as ANNs
Clustering	Unsupervised search for subgroups within the input feature space	<ol style="list-style-type: none"> <li>1. Helpful in finding subgroups when the labels for the larger groups are already known</li> <li>2. Simple and fast</li> </ol>	<ol style="list-style-type: none"> <li>1. Scale- and initialization-dependent</li> <li>2. It's hard to put a number on the variety of subgroups</li> </ol>

## 2 Literature of Review

This section describes the previous work reviews of the Image Based Cardiac Diagnosis Using ML and DL Techniques.

Abubaker et al. (2023) [24] employed the prowess of DL to the public ECG image collection of cardiac patients to expect the four main cardiac abnormalities: history of myocardial infarction, myocardial infarction, abnormal heartbeat, and normal person classes. First, it looked into the transfer learning method; second, it planned a new CNN architecture; and third, it used the aforementioned pre-trained models alongside its own CNN model. These models included the SVM, K Nearest Neighbours (KNN), RF and Decision Tree (DT). The results of the experiments reveal that the proposed CNN model has better precision, recall, accuracy, and F1 score than the state-of-the-art. When using

the NB technique for feature extraction, the recommended CNN model achieves the highest possible score.

Mhamdi et al. (2022) [25] utilized a novel automated deep-learning model for interpreting 12-lead ECG images to diagnose, classify, and predict cardiac arrhythmias. This effort would directly result in better, more cost-effective medical treatment and the prevention of unnecessary deaths. This study provided real-time monitoring using smart mobile technologies simply and cost-effectively (smartwatches, mobile phones, connected T-shirts, etc.). The findings indicated a marginal drop in precision after being implemented on a Raspberry Pi.

Konstantonis et al. (2022) [26] introduced a new ML paradigm for identifying people at moderate to high risk for cardiovascular disease. The efficiency of the ML classifiers was monitored and recorded. The cohort employed a comprehensive set of 46 CVD risk factors, referred to as covariates, within an online cardiovascular framework. The framework's computational efficiency enabled the calculation of these risk factors for each patient in less than 1 s. Compared to the traditional CVD risk score, the cardiovascular framework showed considerable performance improvement. The prediction of CVD in those at medium to high risk has been improved by the ML paradigm.

Khan et al. (2021) [27] focused on calculating the accuracy rate of recognizing the four most common cardiac anomalies. This research uses a dataset that is uncommon for the same kind of work: 11,148 conventional 12-lead-based ECG images were obtained from hospitals and interpreted by specialists in the field. Results showed a strong ability to distinguish between and identify four main cardiac anomalies in the research. The suggested system's accuracy was manually tested by many cardiologists, and they all agreed that it could be used to screen for heart disorders.

Apostolopoulos et al. (2021) [28] explored the efficacy of ML and DL in replacing human clinicians in diagnosis accuracy. It is currently unrealistic to expect even specialists to accurately identify coronary artery disease, but the first stage in developing a state-of-the-art Computer-Aided Diagnostic system is to develop AI models that can compete with the human eye and human expertise. A hybrid multi-input network comprised of RF and InceptionV3 is a promising technique for competing with the accuracy of a medical expert, according to the findings.

Sharifrazi et al. (2021) [29] optimized the process of diagnosing Hypertrophic cardiomyopathy (HCM) by using a DL method. The authors employed a deep CNN to label the augmented images. This was the first time CNN has been used to make a diagnosis of HCM, as far as the team is aware. The method outperforms human judgement on the original dataset, with recall of 97.90%, accuracy of 95.23%, and specificity of 93.06%. To further analyse the suggested strategy, authors have also tested with a variety of optimizers (including Adagrad and Adadelta) and additional data augmentation techniques (including height shift and rotation). Using the data augmentation technique, authors were able to reach an accuracy, which is better than the best accuracy produced by any of the previous data augmentation techniques that have been studied.

Rath et al. (2021) [30] stated an ensemble model that outperforms the single DL model used in this paper by using a Generative Adversarial Network (GAN) and long short-term memory (LSTM). The proposed CNN model outperforms state-of-the-art



methods in terms of accuracy (98.23%), recall (98.22%), F1 score (98.21%), and precision (98.31%).The GAN model is the most effective of the five examined models, whereas the NB model has the lowest detection potentiality.

Moreno et al. (2019) [31] provided a pathology recognition system for cine- MRI data based on a regional multiscale motion interpretation of cardiac structures. A motion description is then built from the occurrences of a collection of flow orientations across all of the multiscale areas. A prediction of the heart condition is obtained by mapping this descriptor to an RF classifier that has been pre-trained. The suggested method was tested on 45 cine-MRI volumes, with an average F1-score achieved when used for the binary classification of four heart diseases.

Table 3 presents a comprehensive summary of the findings derived from the Literature Review.

**Table 3.** Comparative Analysis of LOR.

Author	Techniques Used	Outcomes
Abubaker et al., (2023) [24]	CNN	The proposed CNN model achieves performance metrics that surpass those of previous studies. Specifically, it demonstrates an accuracy of 98.23%, a recall of 98.22%, an F1 score of 98.21%, and a precision of 98.31%
Mhamdi et al., (2022) [25]	CNN	Excellent results were produced by trained models using the architecture that was suggested, with an accuracy of more than 95.00%
Konstantonis et al., (2022) [26]	ML	The CVD-CR system demonstrated a mean accuracy of 98.40% and an AUC of 0.98 ( $p < 0.0001$ ), while the CVD-3YFU system achieved 98.39% and 0.98 ( $p < 0.0001$ )
Khan et al., (2021) [27]	Single Shoot Detection (SSD) Mobile-Net v2-based DNN	The research demonstrated exceptional results in discriminating and identifying four main cardiac anomalies, with an average accuracy of 98%

*(continued)*

**Table 3.** (continued)

Author	Techniques Used	Outcomes
Apostolopoulos et al., (2021) [28]	ML and DL	This approach achieves the same level of accuracy as the expert, which is 79.15% in the specific dataset
Sharifrazi et al., (2021) [29]	Deep CNN	The achieved accuracy of 98.53% surpasses the highest accuracy of 95.83% obtained from other evaluated data augmentation techniques
Rath et al., (2021) [30]	LSTM and GAN	The GAN-LSTM model that was proposed exhibited superior performance in terms of F1-score (0.987), accuracy (0.992), and AUC (0.984) when compared to the previous models
Moreno et al., (2019) [31]	RF	The proposed method was evaluated on 45 cine-MRI volumes, achieving an average F1-score of 77.83% for the binary classification of four cardiac diseases.

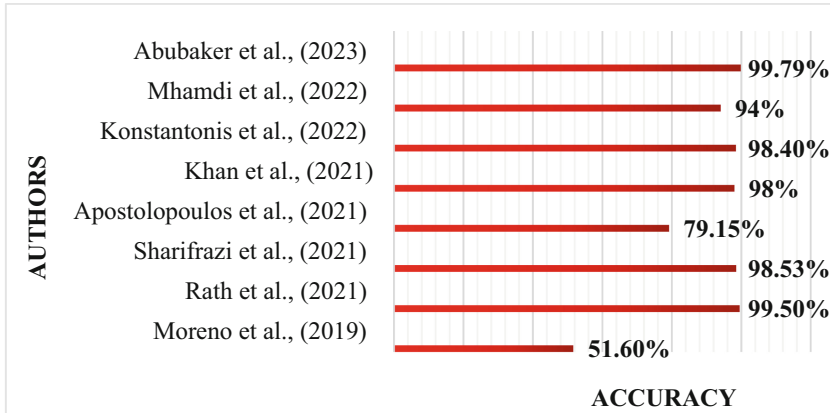
### 3 Comparative Analysis of Reviewing Literatures

This part of the review provides a comparative analysis of the literature that was examined above. By using a variety of datasets, it determined the accuracy that had been found in all of the prior reviewed research papers. After comparative analysis CNN has higher accuracy (99.7%). The LSTM and GAN schemes exhibit the second and third highest levels of accuracy, achieving rates of 99.5% and 98.4% respectively. The RF algorithm exhibits the lowest accuracy at 51.6%. Table 4 presents the results of the comparative analysis conducted on Reviewing Literatures, while Fig. 5 displays the corresponding graph illustrating the comparative analysis.

Figure 5 presents a diagrammatic representation of a comparative study conducted on the data presented in Table 4.

**Table 4.** Compression of Reviewing Literatures

Authors	Techniques used	Dataset used	Accuracy
Abubaker et al.,[24]	CNN	Public ECG images dataset of cardiac patients	99.79%
Mhamdi et al., [25]	CNN	Training and Test dataset	94% and 90%
Konstantonis et al., [26]	ML	SMOTE 5X dataset	98.40%
Khan et al., [27]	SSD Mobile-Net v2-based DNN	Standard 12-lead-based ECG images	98%
Apostolopoulos et al., [28]	ML and DL	Large-scale patient dataset	79.15%
Sharifrazi et al., [28]	Deep CNN	Cine MRI dataset	98.53%
Rath et al., [30]	LSTM and GAN	PTB-ECG dataset and MIT-BIH dataset	99.5% and 98.4%
Moreno et al., [31]	RF	Cardiac cine-MRI	51.6%



**Fig. 5.** Diagrammatic Representation of a Comparative Study

## 4 Conclusion and Future Scope

Diseases of the heart might cause potentially life-threatening damage to the organ and even result in death. To reduce the number of fatalities caused by these conditions, a prompt and precise diagnosis is necessary. The finding of this current study is to ascertain the optimal methodology for predicting CVD, with particular emphasis on the utilization of indicators of cardiac pathology and ECG images. The authors conducted an analysis of various ML and DL models that utilize both image and ECG signal data

for the prediction of CVD. Based on a comparative analysis, it has been determined that CNN exhibits a higher level of accuracy, specifically at 99.7%. The LSTM and GAN models demonstrate the second and third highest levels of accuracy, attaining rates of 99.5% and 98.4% correspondingly. The accuracy of the RF algorithm is observed to be the lowest, measuring at 51.6%. The main purpose of this paper is to examine a wide range of research endeavours that aim to advance the field of cardiovascular medicine. The ultimate aim is to facilitate the development of improved diagnostic techniques and precision medicine approaches specifically within the realm of cardiology.

## References

1. Al-Badi, A., Khan, A.I.: A sustainable development neural network model for Big Data in smart cities. *Procedia Computer Science* **202**, 408–413 (2022). <https://doi.org/10.1016/j.procs.2022.04.057>
2. Bhowmik, T., Bhadwaj, A., Kumar, A., Bhushan, B.: Machine learning and deep learning models for privacy management and data analysis in smart cities. In: Balas, V.E., Solanki, V.K., Kumar, R. (eds.) *Recent Advances in Internet of Things and Machine Learning*. Intelligent Systems Reference Library, vol 215. Springer, Cham (2022). [https://doi.org/10.1007/978-3-030-90119-6\\_13](https://doi.org/10.1007/978-3-030-90119-6_13)
3. Rawat, P., Bajaj, M., Mehta, S., Sharma, V., Vats, S.: A study on cervical cancer prediction using various machine learning approaches. In: *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)* (2023). <https://doi.org/10.1109/icidca56705.2023.10099493>
4. Varshney, M., Bhushan, B., Haque, A.K.M.B.: Big Data Analytics and Data Mining for Healthcare Informatics (HCI). In: Kumar, R., Sharma, R., Pattnaik, P.K. (eds.) *Multimedia Technologies in the Internet of Things Environment*, Volume 3. *Studies in Big Data*, vol 108. Springer, Singapore (2022). [https://doi.org/10.1007/978-981-19-0924-5\\_11](https://doi.org/10.1007/978-981-19-0924-5_11)
5. Predicting chronic kidney disease using ML algorithms and XAIV. *Medicon Engineering Themes* (2023). <https://doi.org/10.55162/mcet.04.142>
6. Pandey, A., et al.: An effective machine learning based heart disease diagnosis analysis. In: *2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN)* (2023). <https://doi.org/10.1109/cictn57981.2023.10141082>
7. Sharma, N.: Multichannel EEG pattern acknowledgment feature assortment using T-test ranking and principal component analysis. In: *2023 1st International Conference on Innovations in High Speed Communication and Signal Processing (IHCSPP)* (2023). <https://doi.org/10.1109/ihcsp56702.2023.10127161>
8. Chae, M., Lee, H.: A prediction of in-hospital cardiac arrest risk scoring based on machine learning. *Int. J. Adva. Sci. Eng. Info. Technol.* **13**(3), 895 (2023). <https://doi.org/10.18517/ijaseit.13.3.17343>
9. Dimension reduction and multiscale modeling for Thin Structures. *Introduction to Multiscale Mathematical Modeling*, 101–123 (2022). [https://doi.org/10.1142/9781800612327\\_0004](https://doi.org/10.1142/9781800612327_0004)
10. Chen, C., et al. Deep Learning for Cardiac Image Segmentation: A Review. *Frontiers in Cardiovascular Medicine*, 7 (2020). <https://doi.org/10.3389/fcvm.2020.00025>
11. Kanani, P., Padole, M.C.: ECG image classification using Deep Learning Approach. *Handbook of Research on Disease Prediction Through Data Analytics and Machine Learning*, 343–357 (2021). <https://doi.org/10.4018/978-1-7998-2742-9.ch016>
12. Sidey-Gibbons, J.A., Sidey-Gibbons, C.J.: Machine learning in medicine: a practical introduction. *BMC Medical Research Methodology* **19**(1) (2019). <https://doi.org/10.1186/s12874-019-0681-4>

13. Zhao, Q., Liu, Z., Adeli, E., Pohl, K.M.: Longitudinal self-supervised learning. *Med. Image Anal.* **71**, 102051 (2021). <https://doi.org/10.1016/j.media.2021.102051>
14. Song, X., Liu, X., Liu, F., Wang, C.: Comparison of machine learning and logistic regression models in predicting acute kidney injury: a systematic review and meta-analysis. *Int. J. Med. Informatics* **151**, 104484 (2021). <https://doi.org/10.1016/j.ijmedinf.2021.104484>
15. Habebh, H., Gohel, S.: Machine learning in Healthcare. *Curr. Genomics* **22**(4), 291–300 (2021). <https://doi.org/10.2174/1389202922666210705124359>
16. Waring, J., Lindvall, C., Umeton, R.: Automated machine learning: review of the state-of-the-art and opportunities for healthcare. *Artif. Intell. Med.* **104**, 101822 (2020). <https://doi.org/10.1016/j.artmed.2020.101822>
17. Azad, C., Bhushan, B., Sharma, R., Shankar, A., Singh, K.K., Khamparia, A.: Prediction model using SMOTE, genetic algorithm and decision tree (PMSGD) for classification of diabetes mellitus. *Multimedia Syst.* (2021). <https://doi.org/10.1007/s00530-021-00817-2>
18. Khamparia, A., et al.: An internet of health things-driven deep learning framework for detection and classification of skin cancer using transfer learning. *Transactions on Emerging Telecommunications Technologies* (2020). <https://doi.org/10.1002/ett.3963>
19. Neural networks referees in 2022. *Neural Networks* **157**, xii–xl (2023). [https://doi.org/10.1016/s0893-6080\(22\)00489-0](https://doi.org/10.1016/s0893-6080(22)00489-0)
20. Qamar, U., Raza, M.S.: Frequent Pattern Mining. *Data Science Concepts and Techniques with Applications*, 271–311 (2023). [https://doi.org/10.1007/978-3-031-17442-1\\_9](https://doi.org/10.1007/978-3-031-17442-1_9)
21. Sarker, I.H., et al.: Cybersecurity Data Science: An overview from machine learning perspective. *Journal of Big Data* **7**(1) (2020). <https://doi.org/10.1186/s40537-020-00318-5>
22. Chopra, D., Khurana, R.: Introduction to machine learning. *Introduction to Machine Learning with Python*, 15–29 (2023). <https://doi.org/10.2174/9789815124422123010004>
23. Bueff, A., Belle, V.: Logic + Reinforcement Learning + deep learning: A survey. In: *Proceedings of the 15th International Conference on Agents and Artificial Intelligence* (2023). <https://doi.org/10.5220/0011746300003393>
24. Abubaker, M.B., Babayiğit, B.: Detection of cardiovascular diseases in ECG images using machine learning and deep learning methods. *IEEE Trans. Artif. Intell.* **4**(2), 373–382 (2023). <https://doi.org/10.1109/tai.2022.3159505>
25. Mhamdi, L., Dammak, O., Cottin, F., Dhaou, I.B.: Artificial Intelligence for cardiac diseases diagnosis and prediction using ECG images on embedded systems. *Biomedicines* **10**(8), 2013 (2022). <https://doi.org/10.3390/biomedicines10082013>
26. Konstantonis, G., et al.: Cardiovascular disease detection using machine learning and carotid/femoral arterial imaging frameworks in rheumatoid arthritis patients. *Rheumatol. Int.* **42**(2), 215–239 (2022). <https://doi.org/10.1007/s00296-021-05062-4>
27. Khan, A.H., Hussain, M., Malik, M.K.: Cardiac disorder classification by electrocardiogram sensing using deep neural network. *Complexity* **2021**, 1–8 (2021). <https://doi.org/10.1155/2021/5512243>
28. Apostolopoulos, I.D., Apostolopoulos, D.I., Spyridonidis, T.I., Papathanasiou, N.D., Panayiotakis, G.S.: Multi-input deep learning approach for cardiovascular disease diagnosis using myocardial perfusion imaging and clinical data. *Physica Med.* **84**, 168–177 (2021). <https://doi.org/10.1016/j.ejmp.2021.04.011>
29. Sharifrazi, D., et al.: Hypertrophic cardiomyopathy diagnosis based on cardiovascular magnetic resonance using Deep Learning Techniques. *SSRN Electronic Journal* (2021). <https://doi.org/10.2139/ssrn.3855445>

30. Rath, A., Mishra, D., Panda, G., Satapathy, S.C.: Heart disease detection using deep learning methods from imbalanced ECG samples. *Biomed. Signal Process. Control* **68**, 102820 (2021). <https://doi.org/10.1016/j.bspc.2021.102820>
31. Moreno, A., Rodriguez, J., Martinez, F.: Regional multiscale motion representation for cardiac disease prediction. 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA) (2019). <https://doi.org/10.1109/stsiva.2019.8730231>



# Short Message Service Spam Detection System for Securing Mobile Text Communication Based on Machine Learning

Ayasha Malik<sup>1</sup>, Veena Parihar<sup>2</sup>, Bharat Bhushan<sup>3</sup>, Alaa Ali Hameed<sup>4</sup>, Akhtar Jamil<sup>5</sup>✉, and Pronaya Bhattacharya<sup>6</sup>

<sup>1</sup> Delhi Technical Campus (DTC), GGSIPU, Greater Noida, India

<sup>2</sup> KIET Group of Institutions, Delhi-NCR, Ghaziabad, India

<sup>3</sup> School of Engineering and Technology (SET), Sharda University, Greater Noida, India

<sup>4</sup> Department of Computer Engineering, Istinye University, Istanbul, Turkey

alaa.hameed@istinye.edu.tr

<sup>5</sup> Department of Computer Science, National University of Computer and Engineering, Islamabad, Pakistan

akhtar.jamil@nu.edu.pk

<sup>6</sup> Department of Computer Science and Engineering, Amity School of Engineering and Technology, Research and Innovation Cell, Amity University, Kolkata, India

pbbhattacharya@kol.amity.edu

**Abstract.** In recent years, the popularity of Short Message Service (SMS) on mobile phones has surged due to technological advancements and the growing prevalence of content-based advertising. However, this has also led to a surge in spam SMS, which can inundate one's handset at any time and potentially result in the theft of personal information. Researchers have explored a range of options to combat spam SMS, including content-based machine learning algorithms and stylometric approaches. While filtering spam emails has proven to be effective, detecting spam SMS presents a unique set of challenges due to the presence of idioms, abbreviations, and well-known terms and phrases that are frequently used in legitimate messages. This study aims to examine and assess different classification techniques by utilizing datasets gathered from previous research studies. The primary emphasis is on comparing conventional Machine Learning (ML) methods. This study specifically investigates the efficacy of classification algorithms, including Logistic Regression (LR), Naïve Bayes (NB), and Random Forest (RF), in accurately identifying spam SMS messages. The results of our research demonstrate that the RF classifier outperforms other traditional ML techniques in the detection of spam SMS messages. This paper thoroughly examines diverse classification techniques employed in spam detection. The knowledge derived from this study has the potential to advance the development of more effective systems for detecting spam SMS messages. Ultimately, this would enhance the security and privacy of mobile phone users.

**Keywords:** Spam detection · SMS spam detection · Machine learning · logistic regression · Text classification

## 1 Introduction

SMS is a widely used communication tool by various enterprises, banks, and government agencies to connect with their clients or customers. In addition, several businesses employ this service for promotional purposes. The ease of access and absence of active internet connection requirements make SMS a crucial architecture. However, with the growing popularity of SMS, it has also become a popular target for spammers and hackers [1]. These individuals can compromise mobile devices by sending malicious links or messages to end-users causing them to release confidential information [2]. As a result, it is essential to limit the content that end-users receive and to develop a system to determine whether a message is SPAM or not, with non-spam messages referred to as HAM. It is relatively easy for a hacker to compromise anyone's cell phone by simply bypassing or transmitting a malicious link to the end-user, the mobile device will be compromised automatically if the end-user clicks on the link or message being transmitted by the hacker/spammer, and we can learn the rest of how a hacker can exploit the system if he gains control of the system. As a result, it has become critical to limit the content that the end-user receives. Consequently, there is a need for a system that can provide the end user with information regarding whether a received message is classified as SPAM or HAM (non-spam). To facilitate research in this area, the SMS Spam Collection has been created. This collection comprises 5,574 English SMS messages that have been categorized as either ham (real) or spam. Each message is represented as a single line in the files, with two columns: v1 indicating the label (ham or spam) and v2 containing the actual content of the message [3]. ML is a fascinating field that integrates parts of various disciplines, including statistics, artificial intelligence theory, data analytics, and numerical techniques. ML is the semi-automated extraction of data [4].

Data sets or data play a crucial role in generating knowledge. This definition can be categorized into three key sections: i) Firstly, when it comes to ML, data serves as the foundation, enabling the extraction of valuable insights and knowledge from the given data set [5]. ii) Secondly, ML emphasizes the need for automation rather than manual extraction of insights from data. It involves leveraging algorithms and techniques to automate the process and derive meaningful information efficiently. Iii) Lastly, while ML heavily relies on automation, it still necessitates human involvement for making intelligent decisions throughout the process to ensure its success. Human participation is vital for handling various aspects and making informed choices during the ML workflow. Simply described, ML is an application that improves its prediction outcomes over time through iterations or experience. Training, as the name implies, is the process by which an application improves with use. Significant iterations may be required to gradually improve results. Data is delivered into an ML algorithm throughout the training process, which refines its internal representation and numerical parameters as deviations or training errors occur [6]. The goal of this stage is to alter the algorithm's internal weights to reduce the cost function and error function or maximize the likelihood. This is referred to as learning when the algorithm's accuracy improves. After the findings, also known as scoring, are accurate enough, the ML application can be utilized to fix the problem [7].



The primary goal of this study is to compare and evaluate different classification techniques and determine the most effective approach for detecting spam SMS. The research study employs traditional ML methods and deep learning techniques to identify the best approach. The results obtained from this study offer valuable insights into the efficiency of different classification techniques and propose potential solutions for minimizing the impact of spam SMS messages. The remainder of the paper is structured as follows: Sect. 2 examines the relevant previous research conducted in this field. It discusses the various papers published in the past and presents a thorough study. Furthermore, Sect. 3 presents the proposed work with a thorough discussion of all the steps included in the work done, along with the description of various ML classification algorithms. It represents the proposed methodology followed for the work. Section 4 presents the outcomes obtained from the proposed work by applying the different ML classifiers on the dataset and also compares the various applied classifiers. Then Sect. 5 concludes the paper by discussing the overall work done and also presents the limitations of the current work and possible future aspects.

## 2 Literature Review

A comprehensive analysis of 18 research papers on SMS spam, Email spam, and Web-spam was conducted to achieve improved accuracy in our project. The survey commenced with a focus on topics related to ML and deep learning approaches. The search process involved utilizing the Google search engine and conducting searches in digital libraries like Google Scholar, Science Direct, IEEE Access, and Research Gate. The search was conducted using keywords such as “Spam Detection for Machine Learning Techniques” and “Deep Learning Techniques” to gather relevant information for this comprehensive review. Research papers were mostly picked relevant to ML algorithms for spam message identification from 2017 to 2021 (Fig. 1, Table 1).

**Table 1.** Comparative Analysis and Gap Study

References	ML models	Summary	Accuracy	Key
Luckner et al. (2019) [3]	SVM and RF	Data from the previous period is used to train a new classifier. They have demonstrated that this method outperforms learning set incrementation. By minimizing spam messages, the system resolves the initial-issues of a lack of learning sets	SVM: 96.2%, RF: 98%	[P3]

(continued)

**Table 1.** (continued)

References	ML models	Summary	Accuracy	Key
Qahtani and Alghazzawi (2019) [7]	DT, SVM, LR, NB, AdaBoost, ANN	In this review, the researchers utilized recall, precision, accuracy, and the CAP Curve for evaluating and comparing the performance of various approaches	DT: 96%, SVM: 94%, LR: 95%, NB: 96.4% RF: 96.5%, AdaBoost: 89% ANN: 99.19%	[P7]
Sjarif et al. (2019) [8]	RF and TF-IDF	The combination of RF and Term Frequency-Inverse Document Frequency was used to detect SMS spam with performance measures including accuracy, precision, and F-measure	RF and TF-IDF: 97.50%	[P8]
Mansoor and Shaker (2019) [9]	NB, CNN	The proposed system employs two classifiers: Naive Bayes (NB), and neural networks. The NB classifier is utilized to classify incoming messages. The neural network will determine whether the incoming message is spam or not	NB: 97%, CNN: 93.9%	[P9]
Mallam Pati et al. (2019) [10]	SVM	In this paper, SVM uses a Kernel function to separate nonlinear data, which is an advanced ML technique in R that improves model accuracy	SVM: 92.4	[P10]
Budi Santoso (2019)	LR, DT, RF	The efficiency of proposed ML algorithms is evaluated based on factors such as the speed of the training and testing processes, as well as the accuracy of detecting spam emails	LR:70%, DT:94% RF:98%	[P2]

(continued)

**Table 1.** (continued)

References	ML models	Summary	Accuracy	Key
Xia and Chen (2020) [12]	CNN, SVM, NB, HMM	The performance evaluation of the proposed Hidden Markov Model (HMM) technique was conducted on a Chinese SMS spam database. The experimental results show that the HMM technique is language-agnostic, as it exhibits effectiveness in identifying spam in both datasets with a high level of accuracy	CNN: 98%, SVM: 94% NB: 87%, HMM:97.6%	[P12]
Asgher (2020) [13]	RF, RT, Adda Boost, DT, NB	Various classified dataset-gathering methods are used in conjunction with the Weka tool. Random Forest and the random tree gave us 100% accurate results	RF: 91%, RT: 91% NB: 78.08%, DT: 92.03% Adda Boost: 85.05%	[P13]
Gomaa (2020) [14]	CNN, SVM, RF, DT, Random Multimodal Deep Learning (RMDL)	The required features are extruded automatically in the proposed methodologies. Using random multi-model deep learning (red) architecture, a benchmark data set of 5574 records was created with a fantastic accuracy of 99.26%	CNN:98.24%, SVM:86.6% DT:94%, RMDL:99.2%	[P14]
Ghourabi et al. (2020) [15]	CNN, LSTM, SVM, KNN, NBM, DT, LR, RF, AdaBoost, Bagging, Extra Trees, CNN-LSMT	Due to the volume of SMS spam increasing over recent years, new security threats like SMiShing have arisen. To address such issues, two deep learning techniques were proposed on the Arabic and English language	SVM: 97.8% KNN: 90% NBM: 97.8% DT: 96.5% LR: 96.5% RF: 97.8% AdaBoost: 97.2% Bagging: 97.2% Extra Trees: 97.8% CNN: 98.1% LSTM: 98.1% CNN-LSTM: 98.3%	[P15]

(continued)

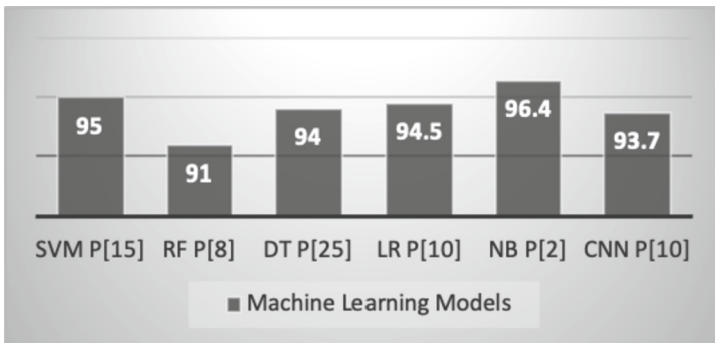
**Table 1.** (continued)

References	ML models	Summary	Accuracy	Key
Hameed (2021) [16]	K-nearest neighbors (KNN)	Based on experimental data, the mutation and crossover variations have outperformed other variations on several performance metrics. Additionally, the mutation and crossover variations have been observed to exceed industry standards	KNN: 96%	[P16]
Liu et al. (2021) [17]	SVM, LR, NB, CNN	The transformer model demonstrates superior accuracy results, showcasing its effectiveness. Additionally, the suggested model exhibits commendable performance on the Twitter dataset	SVM: 81%, LR: 80% NB: 82%, CNN: 80%	[P17]
Gupta et al. (2021) [18]	SVM	To classify SMS messages as spam or ham, the system utilized the TF-IDF vectorizer method. This method was applied to create a vocabulary consisting of all the terms present in spam SMS messages. By analyzing the content of each message and comparing it to the reference vocabulary, the system determines whether the SMS should be classified as spam or ham	SVM: 95%	[P18]
Xia and Chen (2021) [19]	HMM	It improved the performance of HMM models by labeling and assigning weight to words in SMS. This resulted in a formatted observation sequence for the HMM model	HMM: 96.9%	[P19]

(continued)

**Table 1.** (continued)

References	ML models	Summary	Accuracy	Key
Karasoy et al. (2021) [20]	RF, SVM, NB Multinomial (NBM), MLP, LR, KNN, LSTM, CNN, CNN+Word2Vec Random Subspace	Performed content-based classification using 10 ML and deep learning algorithms to filter out SMS spam for the Turkish language	RF: 99.57% NBM: 91.64% SVM: 99.27% MLP: 99.52% Random Subspace: 99.76% LR: 99.64% KNN: 99.34% LSTM: 99.16% CNN: 99.04% CNN+Word2Vec: 99.77%	[P20]
Palimote et al. (2021) [21]	Tensor flow deep learning techniques	The SMS Message dataset, comprising 5574 messages, is categorized using the deep ML model MLPNM	Tensor Flow Deep Learning Techniques: 96.82%	[P21]
Abid et al. (2022) [22]	LR, SVM, RF, GNB, Gradient Boosting (GBM)	Used supervised learning algorithms to perform spam and ham classification of SMS messages. Worked with oversampling and under-sampling techniques	RF: 99% LR: 98% SVM: 97% GBM: 95% GNB: 96%	[P22]
Srinivasarao et al. (2023) [23]	Hybrid SVM and KNN	Performed SMS spam classification and sentiment analysis using a hybrid classifier incorporating SVM and KNN. To further optimize the parameters of the network a Rat Swarm Optimization (RSO) was adopted	Hybrid SVM and KNN: 99.82%	[P23]



**Fig. 1.** Best Accuracy from the papers' analysis

## 3 Types of Spam

### 3.1 A Subsection Sample

SMS spam refers to unwanted or unsolicited text messages that are sent to your mobile phone without your consent, typically for business-related purposes. These messages may consist of a brief statement, a hyperlink to a phone number or text message, a website link for additional details, or a link to download an application. It is important to note that sending unsolicited commercial text messages to a wireless device is prohibited unless the sender has obtained your permission, as outlined by the Federal Trade Commission [16].

### 3.2 Web Spam

The term “web spam” encompasses a range of tactics employed to deceive search engine ranking algorithms, artificially boosting the ranking of search results. These tactics include content spam, which involves generating links to a website to manipulate its link-based score, as well as link spam, which entails creating links to a page to inflate its link-based score. Additionally, cloaking is another strategy where different versions of a page are served to search engine crawlers compared to human users. As internet spam causes frustration to users and disrupts search engines, most commercial search engines actively strive to combat these practices [17].

### 3.3 Email Spam

Unsolicited email spam, commonly referred to as junk email or simply spam, is the term used for unsolicited emails that are sent in large quantities (spamming). These emails are typically sent without prior consent or permission from the recipients and are often part of mass marketing campaigns or fraudulent activities. A Monty Python joke about a canned pork product inspired the name. Spam’s moniker is familiar, inevitable, and tiresome. Email spam has witnessed steady growth since the early 1990s, and by 2014, it was estimated to comprise approximately 90% of all email traffic. Due to the recipient bearing the brunt of the costs, spam can be regarded as unsolicited advertising that effectively imposes a “postage” burden on them [23].

## 4 Proposed Research Work

The use of SMS has grown exponentially in recent years, with technology advancements and content-based advertising being the primary reasons for this increase. However, the rise in SMS usage has also led to an increase in spam messages, which can be not only annoying but also dangerous as they can result in the theft of personal information. To tackle this problem, ML algorithms have been developed to filter spam emails successfully. Recently, researchers have also started using stylistic aspects of text messages to sort them into ham and spam categories. However, detecting spam SMS can still be challenging due to the presence of well-known terms and phrases, idioms, and abbreviations.

The work begins with converting, preparing, and dividing datasets to meet the specific requirements of the algorithms being considered. This involves performing necessary data transformations and preprocessing steps to ensure compatibility and optimal performance with the selected algorithms. The multiple models are then trained, evaluated, and considered for performance evaluation to discover a better approach to spam SMS detection. The following Fig. 2 depicts the flow of the proposed research work.

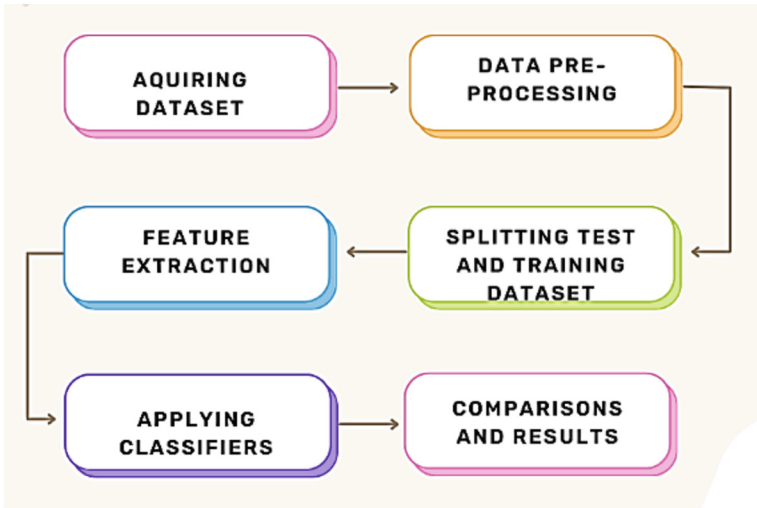


Fig. 2. Flow diagram of the proposed model

#### 4.1 SMS Text (Acquiring Data Set)

To obtain the SMS text dataset, we collected 5573 messages in English, which included both legitimate and spam messages. The dataset was saved in CSV format, where each line represented a text message with two columns: v1 denoting the label (ham or spam) and v2 containing the raw text content. The Spam SMS Dataset 2011–12 was obtained as a compressed file containing multiple text files, each indicating whether its contents were legitimate or spam messages. To consolidate the data, a script was utilized to merge the messages into a single CSV file. Prior to utilizing this data for classification purposes, preprocessing steps will be undertaken to ensure data quality and achieve optimal performance.

## 4.2 Data Pre-Processing

### Using Term Frequency – Inverse Document Frequency (TF-IDF)/Word to Vector Conversion

Text analysis commonly uses the TF-IDF method to extract features from a corpus of documents. TF counts the frequency of a term occurring within a specific document. It measures the number of times a term appears in the document, providing insight into its local importance within that document. IDF, on the other hand, assesses the significance of a term across the entire corpus. It calculates the rarity or importance of a term by considering its frequency across multiple documents. Words that frequently appear in the corpus receive a lower IDF weight, while those that appear less often receive a higher weight. To prepare the text for TF-IDF analysis, several preprocessing steps were carried out. These steps included removing stop words, converting all text to lowercase to eliminate the impact of capitalization, and eliminating non-alphanumeric characters. These actions help in reducing noise and focusing on the meaningful content of the text for accurate TF-IDF analysis. Sklearn's TfidfVectorizer was used to convert the normalized text to a TF-IDF matrix with 5000 features per entry, representing a count vector of words in a bag. This resulting matrix can be used as input to train various ML algorithms for text data classification.

### Using Tokenizer/Text-to-Word Conversion.

Tokenization breaks down text into words or tokens, and keras provides a tokenizer class for this purpose. The tokenizer class in Keras converts text into a sequence of events by vectorizing it. A sequence is a collection of word indices, where the index represents the word's ranking in the dataset. To split the text into words, Keras offers the text-to-word sequence() method. by default, this method separates words by space, converts text to lowercase, and removes punctuation. The number of words considered during tokenization is typically limited to the most frequently occurring terms in the corpus. In our case, the tokenization is limited to 5000 words.

### Training and Test Datasets

The Dataset Was Split into Two Separate Sets, One for Training and One for Testing the Classifiers. The Training Set Consisted of 80% of the Data, While the Remaining 20% Was Reserved for Testing Purposes.

## 4.3 Applying Classifiers

### LR Classifier

LR is an Algorithm Used for Binary Classification, Commonly Employed in Spam Filtering. LR Estimates the Likelihood of a Message Being Classified as Spam or not by Considering Features Such as Word Frequency, Pattern Recognition, and Message Length. It Models the Probability of a Message Belonging to a Particular Class Based on These Features, Enabling Effective Detection of Spam Messages. The Algorithm Learns the Weights Associated with Each Feature, Which Influence the Probability of a Message Being Classified as Spam. These Weights Are Updated During Training to Minimize the Disparity Between the Predicted Probabilities and the Actual Labels of



the Training Data. This is Achieved by Utilizing a Cost Function Such as Cross-Entropy Loss. During the Prediction Phase, the Algorithm Computes the Probability of a New Message Being Spam or not, Using the Learned Weights and the Message's Features. If the Probability Exceeds a Predefined Threshold, the Message is Classified as Spam; Otherwise, It is Classified as Non-Spam [24].

### **NB Classifier**

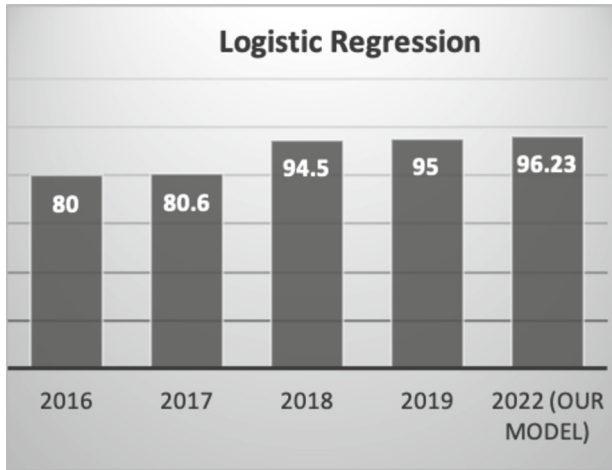
NB is a Probabilistic Classifier that Utilizes Bayes' Theorem to Compute the Probability of a Message Being Spam or not. In Spam Filtering, the Algorithm Determines the Conditional Probability of a Word Appearing in a Spam Message and a Non-Spam Message and Uses These Probabilities to Classify New Messages as Spam or not. The NB Algorithm is a Fast and Efficient Method that Works Well for Text Classification Tasks, Including Spam Filtering [25].

### **RF Classifier**

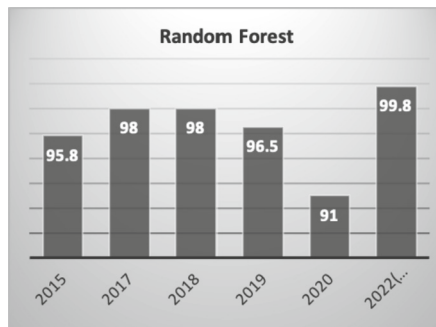
RF is an ensemble learning algorithm well-suited for text classification tasks like spam filtering. It combines multiple decision trees to make predictions, offering robust performance. The algorithm operates by building a considerable number of decision trees, with each tree being trained on a randomly chosen subset of the data. During prediction, RF combines the predictions from all the individual trees to generate a final decision. RF is particularly advantageous for handling high-dimensional data and demonstrates resilience to overfitting. Its ensemble approach enhances the accuracy and generalization capability of the model by leveraging the collective wisdom of multiple decision trees. As a result, RF serves as a powerful tool for effectively addressing text classification challenges, such as spam filtering [26].

## **5 Results Obtained**

This work evaluates and compares various classification techniques on a dataset collected from past research studies. LR was applied as a classifier, which resulted in an accuracy rate of 96.23% on the dataset. This accuracy rate is higher than the previous year's paper's results. NB was also applied as a classifier on the same dataset, and an accuracy rate of 98.22% was obtained, which is higher than the accuracy rate reported in the previous year's paper. Furthermore, the RF classifier was tested on the training dataset, which resulted in the highest accuracy rate of 99.8%. These results suggest that the use of ML techniques can effectively classify text messages into ham and spam categories (Fig. 3, Fig. 4, Fig. 5, Fig. 6).

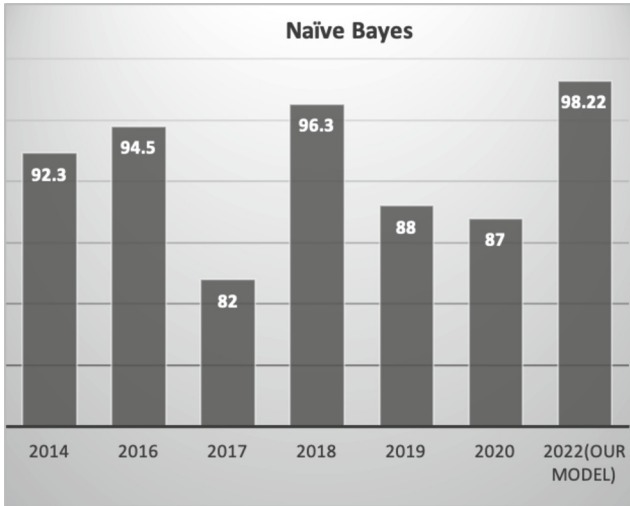


**Fig. 3.** A graphical depiction of the accuracy of last year's Logistic Regression and our current accuracy.

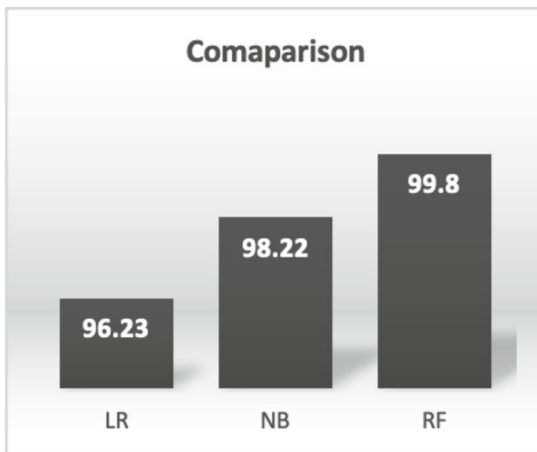


**Fig. 4.** A graphical depiction of the accuracy of last year's Random Forest and our current accuracy.

On applying three classifiers LR, NB, and RF to the data set, the obtained highest accuracy is from RF classifier which is 99.8%. The following Fig. 7 represents the GUI designed to implement the carried-out work.



**Fig. 5.** A graphical depiction of the accuracy of last year’s Naïve Bayes and our current accuracy.



**Fig. 6.** Comparison between the classifiers we used on the data set.

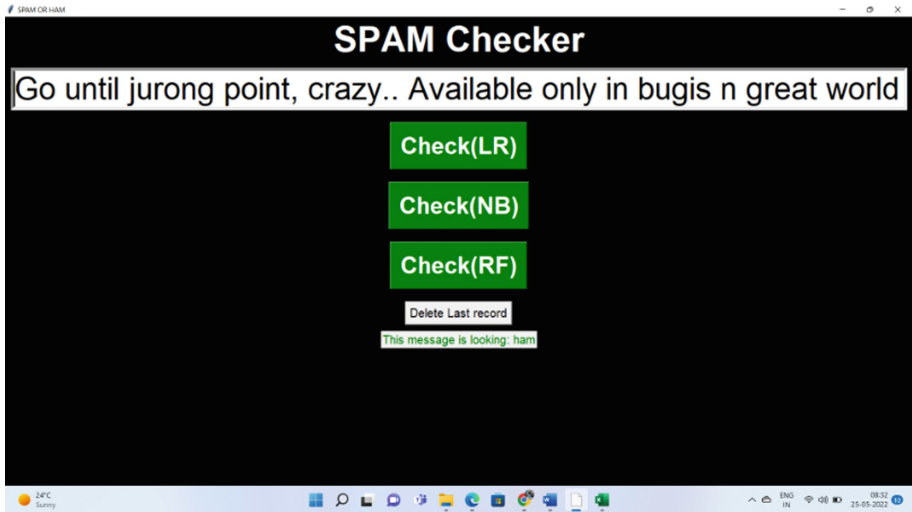


Fig. 7. GUI for the carried-out work

## 6 Conclusion

In conclusion, our research has shown that ML algorithms can effectively distinguish between ham and spam messages with high accuracy. Through our experimentation, the most informative features were identified for classification, such as the presence of certain words or characters, message length, and frequency of specific terms. By using a dataset of labeled SMS messages and implementing various feature extraction techniques, we were able to train and test several models, including NB classifier, RF classifier, and SVMs. Our results showed that RF classifier achieved the highest accuracy of 99.8, indicating better overall performance in terms of precision and recall, followed by Naive Bayes and SVM. We also found that message length and the frequency of specific terms, such as “claim,” “free,” and “prize,” were important predictors for spam messages. These findings demonstrate the effectiveness of ML in detecting spam messages and can be used to develop more accurate and efficient SMS filters. However, there are still some limitations to this research. Our dataset only included messages in English, which may not apply to other languages or cultures. In addition, this study did not account for the evolving nature of spam messages, which may change over time and require updates to the classifier. Future work could focus on improving the performance of the SMS classifier by incorporating more advanced natural language processing techniques, such as sentiment analysis or part-of-speech (POS) tagging, and exploring different ML algorithms, such as neural networks or ensemble models.

## References

1. Bhowmik, T., Bhadwaj, A., Kumar, A., Bhushan, B.: Machine learning and deep learning models for privacy management and data analysis in smart cities. In: Balas, V.E., Solanki,

- V.K., Kumar R. (eds) Recent Advances in Internet of Things and Machine Learning. Intelligent Systems Reference Library, vol. 215. Springer, Cham (2022). [https://doi.org/10.1007/978-3-030-90119-6\\_13](https://doi.org/10.1007/978-3-030-90119-6_13)
2. Soni, S., Bhushan, B.: Use of machine Learning algorithms for designing efficient cyber security solutions. In: 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT) (2019). <https://doi.org/10.1109/icicict46008.2019.8993253>
  3. Luckner, M., Gad, M., Sobkowiak, P.: Antiscam—practical web spam classifier. *Int. J. Electr. Telecommun.* **65**, 713–722 (2019). <https://doi.org/10.24425/ijet.2019.130255>
  4. Sarker, I.H.: Machine learning: algorithms, real-world applications and research directions. *SN Comp. Sci.* **2**(3), 1–21 (2021). <https://doi.org/10.1007/s42979-021-00592-x>
  5. Gaur, J., Goel, A.K., Rose, A., Bhushan, B.: Emerging trends in machine learning. In: 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT). <https://doi.org/10.1109/icicict46008.2019.8993192>
  6. Jindal, M., Gupta, J., Bhushan, B.: Machine learning methods for IoT and their Future Applications. In: International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) (2019). <https://doi.org/10.1109/icccis48478.2019.8974551>
  7. Alqahtani, Sahar & Alghazzawi, Daniyal. (2019). A survey of Emerging Techniques in Detecting SMS Spam. *Transactions on Machine Learning and Artificial Intelligence*. 7. 23–35. <https://doi.org/10.14738/tmlai.75.7116>
  8. Amir Sjarif, N.N., Mohd Azmi, N.F., Chuprat, S., Sarkan, H.M., Yahya, Y., Sam, S.M.: SMS spam message detection using term frequency-inverse document frequency and random forest algorithm. *Proc. Comp. Sci.* **161**, 509–515 (2019). <https://doi.org/10.1016/j.procs.2019.11.150>
  9. Hameed, S., Mansor, H.: Using classification techniques to SMS spam filter. *Int. J. Innov. Technol. Expl. Eng.* **8**, 1734–1739 (2021). <https://doi.org/10.35940/ijitee.L3206.1081219>
  10. Mallampati, D., Shekar, K., Ravikanth, K.: Supervised machine learning classifier for email spam filtering (2019). [https://doi.org/10.1007/978-981-13-7082-3\\_41](https://doi.org/10.1007/978-981-13-7082-3_41)
  11. Santoso, B.: An analysis of spam email detection performance assessment using machine learning. *J. Online Inform.* **4**, 53 (2019). <https://doi.org/10.15575/join.v4i1.298>
  12. Xia, T., Chen, X.: A discrete Hidden Markov Model for SMS spam detection. *Appl. Sci.* **10**, 5011 (2020). <https://doi.org/10.3390/app10145011>
  13. Asgher, T.: A study of SMS spam using machine learning (2020)
  14. Gomaa, W.: The impact of deep learning techniques on SMS spam filtering. *Int. J. Adv. Comp. Sci. Appl.* **11** (2020). <https://doi.org/10.14569/IJACSA.2020.0110167>
  15. Ghourabi, A., Mahmood, M.A., Alzubi, Q.M.: A hybrid CNN-LSTM model for SMS spam detection in Arabic and English messages. *Future Internet* **12**(9), 156 (2020). <https://doi.org/10.3390/fi12090156>
  16. Hameed, S.: Differential evolution detection models for SMS spam. *Int. J. Electr. Comp. Eng.* **11**, 596–601 (2021). <https://doi.org/10.11591/ijece.v11i1.pp596-601>
  17. Liu, X., Lu, H., Nayak, A.: A spam transformer model for SMS spam detection. *IEEE Access*, p. 1 (2021). <https://doi.org/10.1109/ACCESS.2021.3081479>
  18. Gupta, S., Saha, S., Das, S.K.: SMS spam detection using machine learning. *J. Phys: Conf. Ser.* **1797**, 012017 (2021). <https://doi.org/10.1088/1742-6596/1797/1/012017>
  19. Xia, T., Chen, X.: A weighted feature enhanced Hidden Markov Model for spam SMS filtering. *Neurocomputing* **444**, 48–58 (2021). <https://doi.org/10.1016/j.neucom.2021.02.075>
  20. Karasoy, O., Ballı, S.: Spam SMS detection for Turkish language with deep text analysis and deep learning methods. *Arab. J. Sci. Eng.* **47**, 9361–9377 (2022). <https://doi.org/10.1007/s13369-021-06187-1>
  21. Palimote, J., Anireh, V., Nwiabu, N.: A model for filtering spam SMS using deep machine learning technique (2021). <https://doi.org/10.17148/IJARCC.2021.10403>

22. Abid, M.A., Ullah, S., Siddique, M.A., et al.: Spam SMS filtering based on text features and supervised machine learning techniques. *Multimed. Tools Appl.* **81**, 39853–39871 (2022). <https://doi.org/10.1007/s11042-022-12991-0>
23. Srinivasarao, U., Sharaff, A.: Machine intelligence-based hybrid classifier for spam detection and sentiment analysis of SMS messages. *Multimed. Tools Appl.* (2023). <https://doi.org/10.1007/s11042-023-14641-5>
24. Parihar, V., Yadav, S.: Comparative analysis of different machine learning algorithms to predict online shoppers' behaviour. *Int. J. Adv. Netw. Appl.* **13**(6), 5169–5182 (2022). <https://doi.org/10.35444/ijana.2022.13603>
25. Kumar, V., Malik, A.: Heart disease prediction using machine learning. *DTC J. Comput. Intell.* 1(2) (2023). <https://jci.delhitechnicalcampus.ac.in/wp-content/uploads/2022/12/DTC-JCI-4.pdf>
26. Malik, A., Parihar, V., Srivastava, J., Kaur, H., Abidin, S.: Prognosis of diabetes mellitus based on machine learning algorithms. In: 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, pp. 1466–1472 (2023)



# Object Identification: Comprehensive Approach Using Machine Learning Algorithms and Python Tools

Mustafa Al-Asadi<sup>1</sup>✉ and Bharat Bhushan<sup>2</sup>

<sup>1</sup> Faculty of Engineering and Natural Sciences, Computer Engineering Department,  
KTO Karatay University (KTO Karatay Üniversitesi), Konya, Turkey  
mustafa.alasadi@karatay.edu.tr

<sup>2</sup> Department of CSE, School of Engineering and Technology, Sharda University,  
Greater Noida, India

**Abstract.** Computer vision systems have made advancements in object detection using Artificial Intelligence. This paper presents a comprehensive approach utilizing traditional machine learning models such as decision trees, support vector machines, logistic regression, k-nearest neighbors, and naive Bayes. These models are trained on a publicly available dataset from Github, offering diverse objects for recognition. Practical guidelines are provided for easy experimentation. Evaluation metrics include accuracy, precision, recall, and the F1 score. This paper serves as a valuable resource for object identification in the field.

**Keywords:** Object recognition · Machine learning · Python

## 1 Introduction

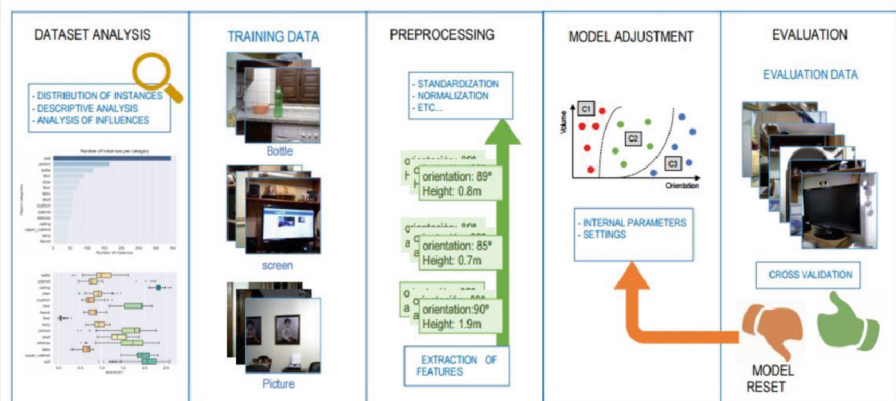
Object recognition systems aim to identify specific categories of elements in a scene from a photograph, such as tables, chairs, or plates. While humans can effortlessly perform this task, implementing it successfully on a computer poses complex algorithmic challenges [1]. Machine Learning (ML) is widely used in computer vision systems to address these challenges (Bishop 2012). ML-based methods adjust their internal parameters using training data to effectively recognize new objects [2–4].

Various techniques and approaches have been proposed for modeling objects in recognition systems. Popular approaches involve extracting features directly from the objects themselves, while others consider their appearance, semantic information, or geometric models (e.g., CAD models). Deep Learning techniques have shown great success in object recognition [5–7]. However, these techniques often require high-performance hardware and extensive training data, which may not always be available [8]. Therefore, this article focuses on ML systems based on characteristic features.

The process of developing object recognition systems can be divided into two main stages: the design stage and the operation stage. The design stage focuses on building and evaluating the machine learning model, while the operation stage involves applying

the model to perform recognition tasks and match objects in a database [9]. Prior to the operation stage, these systems need to go through a comprehensive design phase. Figure 1 illustrates the essential steps involved in this phase, which include dataset analysis, splitting the data into training and testing sets, data preprocessing, method adjustment, and evaluation. With the increasing popularity of these techniques, numerous implementations are now available for testing machine learning methods in object recognition [10–17]. However, it is still relatively uncommon to find well-documented and straightforward solutions that guide the complete design process of a successful system.

In this paper, various libraries were employed for conducting the experiments. The Pandas library [18] was utilized for loading and processing datasets. Seaborn [19] was employed for creating graphical representations. Scikit-learn [11] was used for evaluating recognition methods, including decision trees, support vector machines, logistic regression, k-nearest neighbors, and naive Bayes classification.



**Fig. 1.** Typical steps to follow in the design stage of an object recognition system.

To demonstrate the tasks carried out in each design step, the Robot@Home dataset [20] is utilized. This dataset consists of RGB-D images [21] captured by a mobile robot in various households, providing information on the intensity and depth of the scene. Robot@Home includes labels that assign object categories to different regions within these images [22]. Moreover, it offers a description of each object, encompassing both geometric and visual features.

The key contributions of this study are as follows:

1. The utilization of a publicly available dataset and training traditional machine learning models to address the problem of end-to-end object detection.
2. Providing well-documented and clear solutions to guide the design process of a successful system.



3. Presenting a series of Python scripts available on Github<sup>1</sup> that complement the explanations and analyses presented in this paper.

The remainder of this paper is organized as follows: Sect. 1 provides a brief introduction to the Python libraries used. In Sect. 2, best practices for handling and analyzing the dataset are discussed. Section 3 focuses on object labeling and preparing the necessary descriptions for training the recognition method. Section 4 outlines the adaptation, evaluation, and validation of methods using various techniques and metrics. Finally, Sect. 5 concludes the paper with a brief discussion on its scope.

## 2 Handling and Analysis of the Data Set

The initial phase of designing a successful object recognition system involves studying the dataset, as depicted in Fig. 1. The dataset should consist of instances that belong to the categories targeted for recognition. These instances are typically in the form of observations or extracted features from the objects.

The analysis of the dataset is a crucial step that provides valuable insights into the information it contains [23]. It enables a better understanding of the problem and helps devise a strategy with higher chances of success. Furthermore, it allows for the identification of potential limitations in drawing conclusions during the method evaluation process. By conducting a thorough analysis, possible errors that could lead to incorrect model adjustments can be detected in advance.

To facilitate this analysis, the Pandas library is utilized, as it excels in handling large datasets efficiently and provides the necessary functionalities for extracting valuable information.

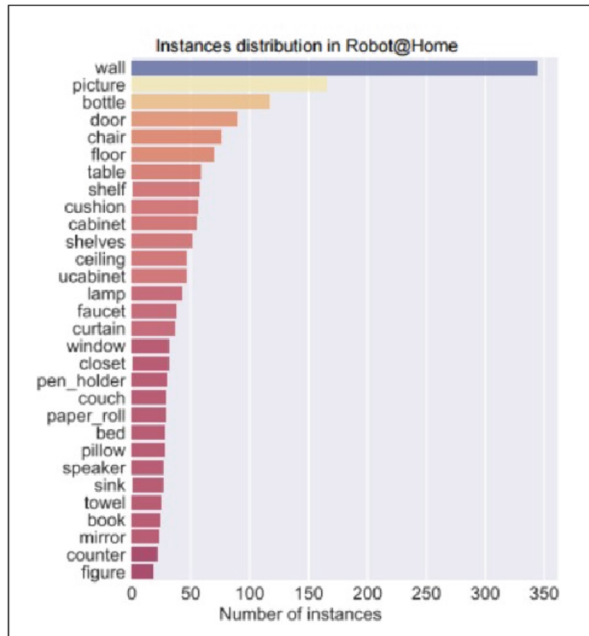
By utilizing Pandas, developers can efficiently load the dataset, examine its structure, and gather relevant information. This analysis phase is instrumental in understanding the dataset's characteristics, such as the distribution of object instances, behavior of descriptive features, and their discriminative capacity. By leveraging Pandas' capabilities, potential errors can be identified early on, leading to a more accurate adjustment of the recognition method.

### 2.1 Distribution of Instances

This analysis aims to determine the balance between the number of instances and categories of objects in the data set. In the case of *Robot@Home*, the results of the analysis show 2,309 cases of objects belonging to 180 classes. However, this number is not real if used to recognize objects, since they appear category with few instances (or even just one). For example, this is the classes' fax, radiator, yoghurt, or yoga ball. Also, as this result is intuited and shown in (Fig. 2), the data set is not balanced, which should be taken into account during the adjustment of the method. Since, according to the one chosen, its performance may be affected. However, if that were the case, the data set can be balanced to solve this problem.

---

<sup>1</sup> [https://github.com/jotaraul/object\\_recognition\\_in\\_python](https://github.com/jotaraul/object_recognition_in_python).



**Fig. 2.** Number of instances of the 30 most repeated object categories in *Robot @ Home*

## 2.2 Descriptive Analysis

The descriptive analysis aims to provide a comprehensive understanding of the different characteristics used to categorize the objects. It involves exploring the central tendency, dispersion, and shape of their distributions. Table 1 presents a basic description of six characteristics in the Robot@Home dataset, computed using the Pandas library.

The information in Table 1 allows for various checks and insights. For example, it can be observed that the minimum height is logically smaller than the maximum height. The average value of the orientation suggests that there are more objects with a vertical orientation in the dataset than a horizontal one. The distribution of object volumes is positively skewed, indicating a predominance of smaller objects, as evident from the percentiles. Additionally, the absence of pure red (corresponding to a tone equal to 0) can be noted based on the minimum and maximum values of the object tones.

This type of information provides verification that the values of the given or computed variables make sense and allows for a deeper understanding of the observed objects.

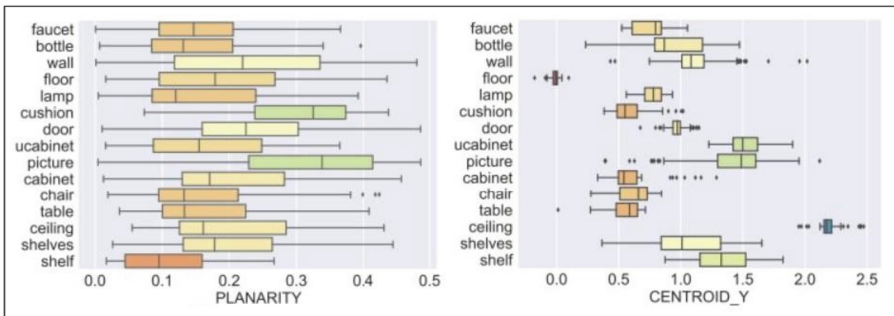
**Table 1.** Basic description of the central tendency, dispersion and shape of the distributions of six characteristics in Robot @ Home

Characteristic	Height min	Max. height	Orientation	Volume	Planarity	Tone
Mean	0.54 m	1.44 m	66.15°	0.80 m <sup>3</sup>	0.21	117.37
STD	0.59 m	0.66 m	32. 33°	1.57 m <sup>3</sup>	0.12	54.50
Minimum	-0.50 m	-0.01 m	0.01°	0.00 m <sup>3</sup>	0.00	24.65
25%	0.00 m	0.87 m	57. 21°	0.03 m <sup>3</sup>	0.11	75.20
50%	0.41 m	1.54 m	84. 76°	0.23 m <sup>3</sup>	0.19	108.95
75%	0.95 m	2.00 m	88. 36°	0.83 m <sup>3</sup>	0.30	154.26
Maximum	2.37 m	2.61 m	89. 99°	13.16 m <sup>3</sup>	0.49	299.10

In addition to numerical descriptions, visual representations can be employed to display the distribution of these characteristics for each category in the dataset. Figure 3 demonstrates the distributions of the planarity and centroid height features using Seaborn tools. The colored boxes represent the interquartile range, with the mean indicated by the vertical line. The vertical exterior lines extend up to 1.5 times the interquartile range and represent potential outliers. The plot helps assess the concentration and differentiation of values across categories.

In the case of planarity, the distribution is considerable, and the values do not show significant differences among categories. Conversely, the centroid height exhibits smaller dispersion and more pronounced differentiation between classes. This suggests that the centroid height feature possesses greater discriminatory power and will contribute more significantly to the recognition process.

Figure 3 provides a visual depiction of the distributions of the planarity and centroid height characteristics, highlighting their variations across different object categories.



**Fig. 3.** Distribution of the characteristics PLANARITY (planarity) and CENTROID Y (height of the centroid from the ground) according to the objects' category

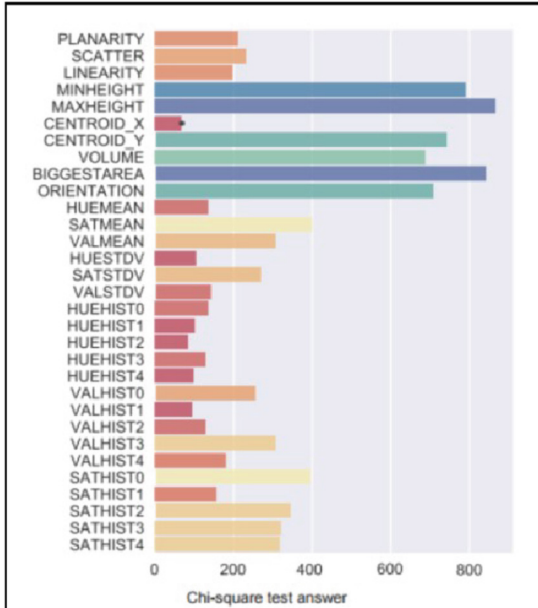
By conducting a descriptive analysis and utilizing Pandas and Seaborn, valuable insights can be gained about the characteristics of objects, facilitating a better understanding of their distributions and aiding in the subsequent steps of the recognition system design.

### 2.3 Influence Analysis

The influence analysis aims to estimate the discriminatory power of the characteristics in relation to the object categories. It determines how promising each characteristic is in differentiating between the types of objects. This analysis is conducted individually since the recognition method will combine these characteristics. A common approach is to perform a chi-square test ( $\chi^2$ ) for each characteristic, which assumes the null hypothesis that there is no relationship between the characteristic and the object categories. The test constructs a model assuming no relationship and evaluates the error between the model and the actual data [24]. The test also provides a p-value, which indicates the statistical significance of the result. A characteristic is considered to have discriminatory power if the chi-square test yields a high value and a low p-value. In this analysis, the characteristics need to be binarized, although it is possible to perform the test with features having more than two values, which would require subsequent post-hoc checks.

Figure 4 presents the results of the chi-square test for all the characteristics in the Robot@Home dataset. The figure displays the test results for each characteristic, with the colors representing the level of significance. The characteristics with a higher value (blue/green color) strongly reject the null hypothesis, indicating a significant relationship between the characteristic and the object categories. There are also variables with a relatively high value (orange color), indicating a moderate level of discrimination. On the other hand, characteristics with lower values (red tones) are less promising in terms of their discriminatory power. In all cases, the associated p-values are considerably lower than the typical threshold of 0.05, which is often used to establish statistical validity for rejecting the null hypothesis.

Figure 4 provides a visual representation of the chi-square test results, highlighting the characteristics' discriminatory power in the Robot@Home dataset.



**Fig. 4.** Results of the chi-square test for all the characteristics in Robot @ Home

### 3 Preprocessing

In the previous section, we have completed a first study of the discriminatory power of the features describing the objects. This step has been possible directly since *Robot@Home* provides, among other information, the objects already related according to the characteristics that appear in Fig. 4. If the data set with which you are working does not provide this information or the result of the study concludes that the characteristics are not appropriate, it would be necessary to complete the data with a varied set of elements and carry out the analyses again. In this sense, and as can be seen in the evaluation section, the collection of Robot@Home features is quite complete, since it describes the size, shape, the spatial position. And the appearance of objects.

As mentioned, the objects can be described from different characteristics, including HOG, SIFT, or SURF, or even Haar characteristics. In some of these cases, the previous analyzes would not be applicable, while in others, they would have to be adapted to make sense and produce valid results.

In any case, automatic learning methods tend to benefit if the characteristics are preprocessed to standardize their values, that is, that they follow a Gaussian distribution of zero mean and deviation. This is due to how they are trained, where if one character has a variance superior to another, it could dominate the model adjustment and compromise its validity. Scikit-learn includes different standardization techniques that even allowed to work with scattered data or with abnormal values. This library also allows, through tools of its preprocessing module, to perform non-linear transformations of the characteristics,

normalize the data (scale each description so that it has a unitary norm), standardize, or even generate polynomial features to add complexity to the model.

## 4 Results and Evaluation of the Method

Once the objects' descriptors have been processed, we are ready to participate in the adjustment phase of the model (see Fig. 1). In this phase, the method's internal parameters are adjusted according to the training data in such a way that given a new description of an object, it can infer its category with success. *Metrics*, once the method has been set, the evaluation data is used to check if its parameters' adjustment is correct. This check is made based on a specific metric, or a set of these, and it is considered that the method is valid when it meets the requirements of success in the particular application. The most popular metrics are accuracy. Accuracy can be defined as the ratio between the number of correctly recognized objects and the number of objects in the dataset. Precision expresses the method's ability not to label an object as belonging to a category that does not belong to his. Recall that measures the capacity to recognize as belonging to a category an object that belongs to it [25].

These metrics can be calculated at the instance (object) level or the object category level, macro. Some metrics combine the previous ones, as it is the F1 score, a weighted average of the recall and the precision. The metric to be used will depend on the requirements that the system must meet.

*Cross-Validation*, on the other hand, divides the data set into training and evaluation subsets are essential. If not done correctly, the validity of conclusions about the success of the method may be compromised. For this, cross-validation techniques are usually used.

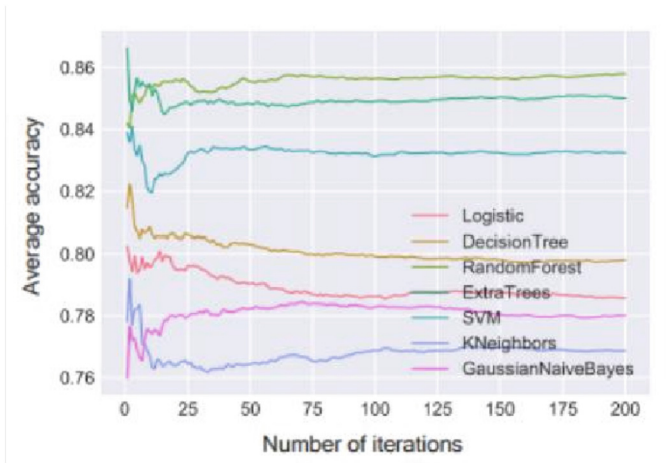
These techniques pursue the validation of a model using a statistical analysis that reinforces the results obtained by giving them a certain degree of generality, limiting the possibility of getting good work due merely to the randomness in the division of data in training and test. It consists of performing an iteration where the data set is divided into two groups, using one for the training and the other for the test. To reduce the variability of the result obtained, this iteration is repeated using different groups, and the validation results are combined to estimate the method's predictive performance.

Different cross-validation techniques define additional ways in which to divide the data and consider a different number of iterations to be performed. In datasets where it is not feasible to check all possible divisions, one of the most used techniques is that of k-fold, wherein each iteration the data are randomly divided into k different groups of the same size, using one of them to evaluate the method and the rest to adjust it. The method's success is obtained after making a high number of iterations (the higher, the more confident we are that the result is valid) and the results obtained in each of them are averaged.

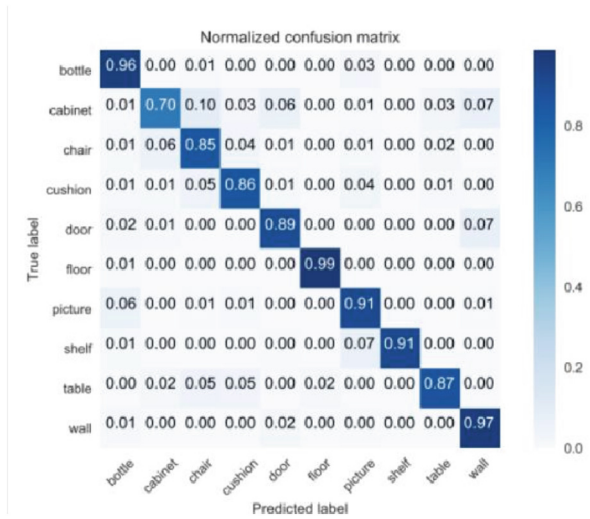
*Methods*, Scikit-learn offers a wide variety of classification methods that can be used for object recognition. These include decision trees, random forests, Extra Trees, support vector machines, logistic regression, k-neighbours, or naive bayesian classifier. This work does not seek to evaluate each one of them scrupulously, but they will be used to illustrate certain facts. For example (Fig. 5) shows how the metric accuracy varies

as a function of the number of completed cross-validation iterations (data obtained using the 15 categories of objects with the highest (n) number of instances and all the available features after standardizing them). That is, the vertical axis shows the average of this metric for the iterations completed so far. You can see how, with a low number of iterations, the result for the same method oscillates considerably, so it cannot be considered reliable. A graph of this type can show us the number of iterations necessary for the metric to stabilize and be informative. In the case that concerns us, we can see how from 100 iterations the values remain stable for most methods.

Each of the methods, apart from their internal parameters to be adjusted during the training, also has a series of configuration parameters that directly influence their performance. They change according to the peculiarities of the application. For the adjustment of these parameters, cross-validation can also be used, adding the configuration parameters to adjust the number of iterations and groups' number to the variables. Confusion matrix, A graphical way to check different aspects of a method's performance is to use the so-called confusion matrix. This matrix indexes the objects' categories in the evaluation data set in its rows, while the column's index the categories to which they have been assigned by the method. Thus, the greater the concentration on the matrix's diagonal, the more successful it would be. This matrix also allows us to observe for which categories of objects the model has lower performance, to be able to readjust the design phase to solve it. In (Fig. 6) shows the confusion matrix using the random forest method, constructed by accumulating the results reported by it in 200 cross-validation iterations (to improve visibility, only the ten categories are shown) with more instances. We can see how the method's behaviour, in general, is quite good (values at one on the diagonal), and that the most conflictive category is closet. The most confusing type is a chair; followed by the wall, so if you want to improve these results, it would be necessary to add some feature that allows you to better discriminate between these categories.

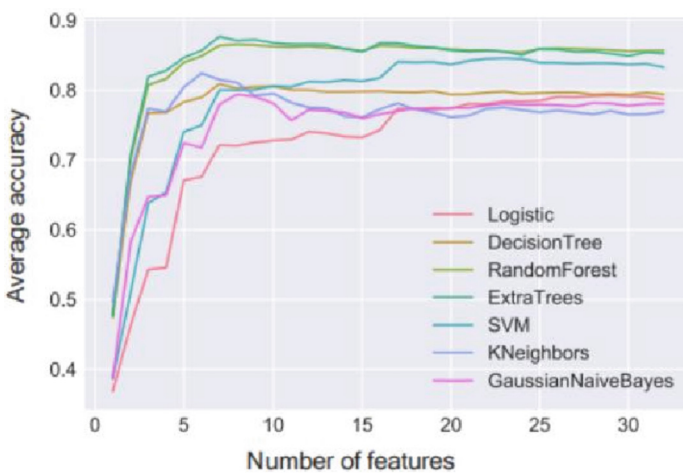


**Fig. 5.** Accuracy of a series of methods for recognition according to the number of cross-validation iterations performed.



**Fig. 6.** Confusion matrix for the ten categories with the highest number of instances using random forests

Features, In Sec. 3, the discriminatory power of the elements in Robot@Home was analyzed. The use of non-discriminatory features affects the methods in different ways: some can isolate them and give them less weight, not affecting their performance, while others suffer from their use. In (Fig. 7) shows the evolution of the accuracy of the methods used in this work, depending on the number of features used (again using 200 cross-validation iterations). This number always refers to the most favourable characteristics, as shown in Fig. 7. Another factor to be taken into account is the time used in adjusting the models, which can be critical according to the application. For example, while the



**Fig. 7.** The accuracy of the methods depends on the number of characteristics used.



adjustment using six features requires a total of 32 s, using the 32 s, it amounts to 72 s. (data were taken with an Intel (R) Core (TM) i7-3820 CPU @ 3.60 GHz and 16 GB of RAM).

## 5 Conclusions and Future Work

This work has provided a series of guidelines for designing methods based on automatic learning capable of recognizing objects to achieve high accuracy. The literature has dealt with several techniques that address the problem of the recognition of objects. However, what is not so frequent is to find straightforward solutions to complete the necessary steps in the design of a successful system [26]. In this study, the text is completed by Python scripts, available at: [https://github.com/jotaraul/object\\_recognition\\_in\\_python](https://github.com/jotaraul/object_recognition_in_python), which allow each of the described steps to be practically tested to meet the design phase using tools from the Pandas, Seaborn and Scikit-learn libraries. We have used the CPU for processing in the project. Future enhancements can be focused on by implementing the project on the system having GPU for faster results and better accuracy. This study will be serving as the guiding thread for the design of successful methods. In addition, we think that the article can be useful in teaching tasks and any enthusiast who wants to build his recognizer.

## References

1. Oliveira, M., et al.: Concurrent learning of visual codebooks and object categories in open-ended domains. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE (2015)
2. Rasheed, J., et al.: A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images. *Interdiscip. Sci. Comp. Life Sci.* **13**, 103–117 (2021)
3. Guney, S., et al.: Abalone age prediction using machine learning. In: Mediterranean Conference on Pattern Recognition and Artificial Intelligence. Springer (2021)
4. Zontul, M.: Customer Credit Rating Estimation Using Machine Learning Methods. *Manchester J. Artif. Intell. Appl. Sci.* **1**(1) (2020)
5. Knopp, J., et al.: Hough transform and 3D SURF for robust three dimensional classification. In: European Conference on Computer Vision. Springer (2010)
6. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger. arXiv preprint (2017)
7. Hussain, S.A., Al Balushi, A.S.A.: A real time face emotion classification and recognition using deep learning model. In: *J. Phys. Conf. Ser.* IOP Publishing (2020)
8. Camilleri, D., Prescott, T.: Analysing the limitations of deep learning for developmental robotics. In: *Conference on Biomimetic and Biohybrid Systems*. Springer (2017)
9. Kim, J.-Y., et al.: A 201.4 GOPS 496 mW real-time multi-object recognition processor with bio-inspired neural perception engine. *IEEE J. Solid-State Circuits* **45**(1), 32–45 (2009)
10. Bradski, G.: The opencv library. *Dr. Dobb's J. Softw. Tools* (2000)
11. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. *J. Machine Learn. Res.* **12**(Oct), 2825–2830 (2011)
12. Ruiz-Sarmiento, J.-R., Galindo, C., Gonzalez-Jimenez, J.: A survey on learning approaches for undirected graphical models. Application to scene object recognition. *Int. J. Approx. Reasoning* **83**, 434–451 (2017)

13. Ruiz-Sarmiento, J.-R., Galindo, C., Gonzalez-Jimenez, J.: Modelado del contexto geométrico para el reconocimiento de objetos. *Actas de las XXXVIII Jornadas de Automática* (2017)
14. Ruiz-Sarmiento, J., et al.: mVision, a toolbox for computer vision courses. In: *The 12th annual International Technology, Education and Development Conference (INTED2018)* (2018)
15. Wang, L., et al.: Object detection combining recognition and segmentation. In: *Asian conference on computer vision*. Springer (2007)
16. Thomas, A., et al.: Towards multi-view object class detection. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. IEEE (2006)
17. Cyganek, B.: *Object Detection and Recognition in Digital Images: Theory and Practice*. John Wiley & Sons (2013)
18. McKinney, W.: Data structures for statistical computing in python. In: *Proceedings of the 9th Python in Science Conference, Austin, TX* (2010)
19. Waskom, M., et al.: Seaborn: statistical data visualization. <https://seaborn.pydata.org/>(visited on 2017-05-15) (2014)
20. Ruiz-Sarmiento, J.R., Galindo, C., Gonzalez-Jimenez, J.: Robot@ home, a robotic dataset for semantic mapping of home environments. *Int. J. Robot. Res.* **36**(2), 131–141 (2017)
21. Ruiz-Sarmiento, J., et al.: Navegación reactiva de un robot móvil usando Kinect. *Universidad de Málaga, Campus de Teatinos, 29071 Málaga* (2011)
22. Ruiz-Sarmiento, J.-R., Galindo, C., Gonzalez-Jimenez, J.: Olt: A toolkit for object labeling applied to robotic RGB-D datasets. In: *Mobile Robots (ECMR), 2015 European Conference on*. IEEE (2015)
23. Al-Asadi, M.: Interval-valued data analysis: a review. *Artif. Intell. Stud.* **5**(2), 47–55 (2022)
24. Al-Asadi, M.A., Tasdemir, S.: Empirical comparisons for combining balancing and feature selection strategies for characterizing football players using FIFA video game system. *IEEE Access* **9**, 149266–149286 (2021)
25. Al-Asadi, M., Taşdemir, Ş., Örnek, H.K.: Predict the number of traffic accidents in turkey by using machine learning techniques and python tools. *Artif. Intell. Stud.* **5**(2), 35–46 (2022)
26. Al-Asadi, M.A., Taşdemir, Ş. Tezcan, B.: An online information system for football club management. In: *Kongre Kitapçığı/Congress Proceedings Book* (2018)

# Author Index

## A

Aattouchi, Issam 68  
Aggarwal, Ashima 354  
Agrawal, Jitendra 324  
Ahmad, Akram 363  
Airen, Sonu 324  
Ait Kerroum, Mounir 68  
Ait Temghart, Abdelkarim 156  
Akhtar, Sohaib 420  
Akshaya, J. 112  
Al Abidin Ibrahim, Zein 386  
Al Mamun, Md. Abdullah 231  
Al-Asadi, Mustafa 508  
Aljeri, Abdullah 241  
Arana, A. 334  
Asri, Hiba 281  
Auccapuri, D. 334  
Ávila, M. 92

## B

Bakirci, Murat 58  
Baldeón, J. 334  
Balta Kaç, Seda 8  
Banerjee, Sheekar 310  
Banu, Ruksana 344  
Baslam, Mohamed 156  
Belangour, Abdessamad 1  
Belkeziz, Radia 404  
Benazzi, Abdelhamid 79, 181  
Benazzi, Naima 79  
Bend, Julia 122  
Bhatnagar, Mohit 166  
Bhattacharya, Pronaya 492  
Bhushan, Bharat 455, 465, 478, 492, 508  
Bourday, Rachid 68  
Bowden, Adam 297  
Büyüktanir, Büşra 23

## C

Caro, A. 92  
Chávez, P. 334

Chefira, Reda 404  
Chowdhury, Anupam Hayath 231

## D

Das, Biva 478  
Dbouk, Mohamed 386  
Dhamini, C. H. 271  
Dhanya, M. 112  
Díaz, E. 334  
Direkoglu, Cem 262  
Dursun, Taner 192

## E

Eken, Süleyman 8, 143  
Ekinci, Ekin 143  
El Koufi, Nouhaila 1  
Eswar Chowdary, D. 112

## F

Farnè, Matteo 213

## G

Gálvez, R. 334  
Ganapathy, Jayanthi 271  
Gmira, Faiq 181  
Gopalakrishnan, E. A. 112, 132

## H

Hameed, Alaa Ali 455, 465, 478, 492  
Harsha, D. 112  
Hasan, Mohammad Kamrul 231  
Hernández, V. 334  
Hossain, Md. Jubayer 231  
Huchhanavar, Shivraj 166

## I

Ibrahim, Abdullahi Abdu 443  
Igbasanmi, Oluwaseyi 297  
Ilyas, Muhammad 432  
Inan, Timur 223

**J**

Jamil, Akhtar 455, 465, 478, 492  
 Jarir, Zahi 281  
 Jarjar, Abdellatif 79, 181  
 Jarjar, Mariem 79, 181  
 Jithin Mohan, K. S. 43  
 Joga, S. Ramana Kumar 251  
 Juyal, Aayush 455, 465  
 Jyothish Lal, G. 132

**K**

Kabir, S. Rayhan 231  
 Kadhim, Ali Salam 386  
 Kanj, Hassan 241  
 Kattass, Mourad 79, 181  
 Khalifa, Tarek 241  
 Korkut, Sila Ovgu 32, 102

**L**

Lau, M. 334

**M**

Malik, Ayasha 492  
 Márquez, Fausto Pedro García 251, 271  
 Marwan, Mbarek 156  
 Masuda, A. 334  
 Merza, Hayder Mosa 386  
 Mezher, Karrar Hamzah 223  
 Mohamed Ali, Arsyia 132  
 Molano, R. 92  
 Monir, Md. Kamrul Hasan 310  
 Morales-Arsenal, Roberto 288

**N**

Naeem, M. Asif 420  
 Nag, Anindya 478

**O**

Onan, Aytug 32  
 Öörni, Anssi 122  
 Özer, Muhammed Mirac 58

**P**

Pandey, Shraiayash 455, 465  
 Parihar, Veena 492  
 Patino, Gustavo 202  
 Paul Sathiyam, S. 43  
 Paul, Kaushik 251  
 Pinar-Pérez, Jesús María 288  
 Pranav Kumaar, B. E. 112

**Q**

Qobbi, Younes 79

**R**

Rahman, Mehtab Ur 262  
 Rahul, G. 112  
 Raisa, Meher Durdana Khan 231  
 Rehman, Abdullah 432  
 Rodríguez, P. G. 92  
 Rrghout, Hicham 79, 181

**S**

Sadek, Islam M. Momtaz A. 443  
 Salman, Ali 420  
 Sancho, J. C. 92  
 Santa, Santiago 202  
 Sbeity, Ihab 386  
 Shabbir, Bilal 420  
 Shagar, Md. Tanvir Miah 231  
 Sharma, Anurag 354  
 Sil, Riya 478  
 Singh, Vaishali 363  
 Sinha, Pampa 251  
 Sirakov, Nikolay M. 297  
 Sowmya, V. 112, 132  
 Suffian, Muhammad 376  
 Sutradhar, Utsa Chandra 231

**T**

Takaoğlu, Faruk 192  
 Takaoğlu, Mustafa 192  
 Tibssirte, Oumaima 404  
 Turan, Salih Can 23

**U**

Ulker, Erman 32  
 Upreti, Kamal 363

**V**

Velez, Diego 202

**Y**

Yalcin, Femin 32  
 Yang, Meng 213  
 Yildiz, Kazim 23  
 Yildiz, Umut 102  
 Yurtsever, M. M. Enes 143

**Z**

Zazpe-Quintana, María Pilar 288