

Telex M. N. Ngatched
Isaac Woungang
Jules-Raymond Tapamo
Serestina Viriri (Eds.)



631

Pan-African Artificial Intelligence and Smart Systems

LNICST

Third Pan-African Conference, PAAISS 2024
Durban, South Africa, December 4–6, 2024
Proceedings, Part I

Part 1



Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering

631

Editorial Board Members

Ozgur Akan, *Middle East Technical University, Ankara, Türkiye*

Paolo Bellavista, *University of Bologna, Bologna, Italy*

Jiannong Cao, *Hong Kong Polytechnic University, Hong Kong, Hong Kong*

Geoffrey Coulson, *Lancaster University, Lancaster, UK*

Falko Dressler, *University of Erlangen, Erlangen, Germany*

Domenico Ferrari, *Università Cattolica Piacenza, Piacenza, Italy*

Mario Gerla, *UCLA, Los Angeles, USA*

Hisashi Kobayashi, *Princeton University, Princeton, USA*

Sergio Palazzo, *University of Catania, Catania, Italy*

Sartaj Sahni, *University of Florida, Gainesville, USA*

Xuemin Shen , *University of Waterloo, Waterloo, Canada*

Mircea Stan, *University of Virginia, Charlottesville, USA*

Xiaohua Jia, *City University of Hong Kong, Kowloon, Hong Kong*

Albert Y. Zomaya, *University of Sydney, Sydney, Australia*

The LNICST series publishes ICST's conferences, symposia and workshops.

LNICST reports state-of-the-art results in areas related to the scope of the Institute.

The type of material published includes

- Proceedings (published in time for the respective event)
- Other edited monographs (such as project reports or invited volumes)

LNICST topics span the following areas:

- General Computer Science
- E-Economy
- E-Medicine
- Knowledge Management
- Multimedia
- Operations, Management and Policy
- Social Informatics
- Systems

Telex M. N. Ngatched · Isaac Woungang ·
Jules-Raymond Tapamo · Serestina Viriri
Editors

Pan-African Artificial Intelligence and Smart Systems

Third Pan-African Conference, PAAISS 2024
Durban, South Africa, December 4–6, 2024
Proceedings, Part I

Editors

Telex M. N. Ngatched 
McMaster University
Hamilton, ON, Canada

Isaac Woungang 
Toronto Metropolitan University
Toronto, ON, Canada

Jules-Raymond Tapamo 
University of KwaZulu-Natal
Durban, South Africa

Serestina Viriri 
University of KwaZulu-Natal
Durban, South Africa

ISSN 1867-8211

ISSN 1867-822X (electronic)

Lecture Notes of the Institute for Computer Sciences, Social Informatics
and Telecommunications Engineering

ISBN 978-3-031-94441-3

ISBN 978-3-031-94442-0 (eBook)

<https://doi.org/10.1007/978-3-031-94442-0>

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2025

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

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

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

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

Preface

We are delighted to introduce the proceedings of the third edition of the Pan-African Artificial Intelligence and Smart Systems Conference (PAAISS 2024). This conference brought together leading academic research scientists, industry practitioners, independent scholars, and innovators from across the African continent and the world, to explore, exchange, and discuss the challenges and opportunities of harnessing the capabilities of Artificial Intelligence (AI) and Smart Systems (SS), which have emerged as the engine of the next wave of future innovations.

The theme of PAAISS 2024 was “Advancing AI research in Africa”. The technical program of PAAISS 2024 consisted of 39 full papers, carefully selected from 103 submissions, for oral presentation sessions at the conference tracks. All the submissions were full papers with no invited papers. The peer review process was conducted with the highest standards of academic rigor and integrity. All submissions underwent a rigorous single-blind peer review process, where each submission was reviewed by a minimum of 3 reviewers, who assessed the contributions based on their originality, scientific merit, relevance to the conference themes, and overall quality. In keeping with our commitment to fairness and transparency, any contributions co-authored by members of the conference committee were handled with particular care. To avoid any potential conflicts of interest, these submissions were reviewed by external reviewers who had no affiliation with the authors or the committee member. This ensured an objective evaluation of all contributions, maintaining the integrity of the peer review process.

Apart from the presentations of high-quality technical papers, the technical program also featured a welcoming address by Neil Koorbanally, acting DVC of Research, University of KwaZulu-Natal, South Africa, an address by Thomas Ndousse-Fetter, Imhotep Geostrategic Innovations, USA, and five keynote speeches. First, we would like to extend our heartfelt thanks to Neil Koorbanally and Thomas Ndousse-Fetter, whose words set the stage for a thought-provoking and inspiring conference, capturing the essence of the themes we explored throughout the event, and challenging us to think critically about the future of AI and SS in Africa. Then, we would like to extend our deepest thanks to our distinguished keynote speakers, whose insightful and stimulating presentations enlightened the conference. Their expertise and vision sparked rich discussions. Their contributions truly were a highlight of this event, and we are incredibly grateful for their participation. The five keynote presenters were Michael Felsberg, Linköping University, Sweden; Bigomokero Antoine Bagula, University of Western Cape, South Africa; Sudeep Tanwar, Nirma University, India; Foutse Khomh, Polytechnique Montréal, Canada, and Jerry John Kponyo, Kwame Nkrumah University of Science and Technology, Ghana.

An international conference of this size requires the support and hard work of many people to produce a successful PAAISS 2024 technical program and conference proceedings. Our Steering Committee chair, Dr. Thomas Ndousse-Fetter was essential for the success of the conference. We sincerely appreciate his vision, constant support, and

guidance. Our sincere appreciation also goes to the excellent Organizing Committee, whose tireless work behind the scenes ensured the success of this conference. From the early stages of planning to the final execution, their commitment to detail, organization, and the overall success of the event has been essential. We are grateful for their vision and dedication in making this conference a reality.

We are equally thankful to the various Track Chairs and Co-chairs, who played a crucial role in curating the content and overseeing the sessions. Their expertise and leadership helped shape the high quality of the program and facilitated engaging discussions across the various themes of the conference.

Our heartfelt appreciation also goes to the Technical Program Committee members, whose meticulous, rigorous, and timely review of the submitted papers ensured that only the highest-quality research was presented at the conference and made it into these proceedings. Their valuable feedback and constructive critiques have significantly elevated the caliber of this year's proceedings, and we are grateful for their time and effort. We also would like to express our sincere thanks to the authors of the papers featured in these proceedings. Your hard work, creativity, and dedication have made this conference a forum for sharing new ideas and advancing the frontiers of AI and SS. Without your contributions, this event would not have been possible, and we are grateful for the opportunity to showcase your research.

Finally, we would like to express our sincere gratitude to our hosts and sponsors, the University of KwaZulu-Natal and the Convention Bureau of Tourism and Film KwaZulu-Natal. Their generous support not only made the conference possible but also allowed us to enhance the overall experience for all participants. Special thanks to Jules-Raymond Tapamo, Serestina Viriri, and the University of KwaZulu-Natal administrators who led the local organization. We are also grateful to the Southern Sun Elangeni Maharani hotel for hosting the conference for in-person sessions.

We strongly believe that the PAAISS conference will continue in the coming years to be an excellent forum for all researchers, developers, and practitioners to discuss the related up-to-date science, technology, and practical aspects of Artificial Intelligence and Smart Systems. We also expect that the future PAAISS conferences will be as successful and stimulating, as shown by the contributions presented in this volume. We do hope that you will enjoy reading these proceedings.

Isaac Woungang
Telex M. N. Ngatched
Jules-Raymond Tapamo
Serestina Viriri
Thomas Ndousse-Fetter

Organization

Steering Committee

Thomas Ndousse-Fetter	Imhotep Research Institute, USA
Isaac Woungang	Toronto Metropolitan University, Canada
Telex M. N. Ngatched	McMaster University, Canada
Jules-Raymond Tapamo	University of KwaZulu-Natal, South Africa
Serestina Viriri	University of KwaZulu-Natal, South Africa

Organizing Committee

General Chair

Thomas Ndousse-Fetter	Imhotep Geostrategic Innovations, USA
-----------------------	---------------------------------------

General Co-chairs

Isaac Woungang	Toronto Metropolitan University, Canada
Serestina Viriri	University of KwaZulu-Natal, South Africa

TPC Chairs and Co-chairs

Telex M. N. Ngatched	McMaster University, Canada
Jules-Raymond Tapamo	University of KwaZulu-Natal, South Africa

Sponsorship and Exhibit Chair

Jules-Raymond Tapamo	University of KwaZulu-Natal, South Africa
----------------------	---

Local Chairs and Co-chairs

Tom Walingo	University of KwaZulu-Natal, South Africa
Jean Vincent Fonou Dombeu	University of KwaZulu-Natal, South Africa

Award Chair

Pius Adewale Owolawi	Tshwane University of Technology, South Africa
----------------------	--

Publicity and Social Media Chairs and Co-chairs

Justice Owusu Agyemang	Kwame Nkrumah University of Science and Technology, Ghana
Paul Python	University Cheikh Anta Diop of Dakar, Senegal

Publications Chair

Telex M. N. Ngatched	McMaster University, Canada
----------------------	-----------------------------

Web Chair

Justice Owusu Agyemang	Kwame Nkrumah University of Science and Technology, Ghana
------------------------	--

Panels Chairs and Co-chairs

Sylvester B. Aboagye	University of Guelph, Canada
Jules M. Moualeu	University of the Witwatersrand, South Africa

Track Co-chairs

Mandla Victor Gwetu	Stellenbosch University, South Africa
Jean Vincent Fonou Dombeu	University of KwaZulu-Natal, South Africa
Edgar Jembere	University of KwaZulu-Natal, South Africa
Pius Adewale Owolawi	Tshwane University of Technology, South Africa
Olutayo O. Oyerinde	University of the Witwatersrand, South Africa
Okuthe Paul Kogeda	University of KwaZulu-Natal, South Africa
Alain Richard Ndjiongue	University of the Witwatersrand, South Africa
Clement Nyieranda	University of the Western Cape, South Africa
Tom Walingo	University of KwaZulu-Natal, South Africa
Hongjun Xu	University of KwaZulu-Natal, South Africa
Jules M. Moualeu	University of the Witwatersrand, South Africa
Sylvester B. Aboagye	University of Guelph, Canada

Technical Program Committee

A. B. M. Bodrul Alam
 Faria Khandaker
 Anish Kurien
 Herve Frackin
 Nico Steyn
 Omar Maraqa
 Kingsley A. Ogudo
 Akintunde Ayodeji Alonge
 Sanjay Kumar Dhurandher
 Dane Brown
 Stones Chindipha
 Vinesh Kumar
 Bhekisizwe Mthethwa
 Yuval Genga
 Hamed Aly
 Megha Gupta
 Oluwaseyi Babalola

Clopas Kwenda
 Khulumani Sibanda
 Weston Govere
 Thambo Nyathi
 Mthulisi Velepini
 Dwane Brown
 Pravesh Ranchod
 Alok Shukla
 Sourabh Solanki
 Deemah Tashman
 Pedro Gorla Silva
 Collins Leke
 Majid H. A. Khoshafa
 Jackson Phiri
 Mehrdad Ghaziasgar
 Pakiso Khomokhoana
 Vitalis Ozianyi
 Patrick Nkwari
 Narushan Pillay
 Frank Igboamalu
 Boriane Tchaleu
 Carl Kugblenu

Algoma University, Brampton Campus, Canada
 Algoma University, Brampton Campus, Canada
 Tshwane University of Technology, South Africa
 University of Le Havre, France
 Tshwane University of Technology, South Africa
 McMaster University, Canada
 University of Johannesburg, South Africa
 University of Johannesburg, South Africa
 University of Delhi, India
 Rhodes University, South Africa
 Rhodes University, South Africa
 University of Delhi, India
 University of KwaZulu-Natal, South Africa
 University of the Witwatersrand, South Africa
 Dalhousie University, Canada
 University of Delhi, India
 Cape Peninsula University of Technology, South Africa
 Africa
 Great Zimbabwe University, Zimbabwe
 Walter Sisulu University, South Africa
 Arrupe Jesuit University, Zimbabwe
 University of Pretoria, South Africa
 University of Limpopo, South Africa
 Rhodes University, South Africa
 University of the Witwatersrand, South Africa
 University of the Witwatersrand, South Africa
 University of Luxembourg, Luxembourg
 Polytechnique Montréal, Canada
 Lappeenranta University of Technology, Finland
 University of Johannesburg, South Africa
 McMaster University, Canada
 University of Zambia, Zambia
 University of the Western Cape, South Africa
 University of the Free State, South Africa
 Strathmore University, Kenya
 University of Toronto, Canada
 University of KwaZulu-Natal, South Africa
 University of Johannesburg, South Africa
 University of Johannesburg, South Africa
 Aalto University, Finland

Mohamed Selim	Memorial University of Newfoundland, Canada
Saba Asaad	York University, Canada
Fanoro Mokesioluwa	University of Johannesburg, South Africa
Sandile Mhlanga	University of Johannesburg, South Africa
Roger Kwao Ahiadormey	Institute for Scientific and Technological Information, Council for Scientific and Industrial Research, Ghana
Elie Ngomseu Mambou	John Abbott College, Canada
Lawrence Githuari	University of Nairobi, Kenya
Khulumani Sibanda	University of Fort Hare, South Africa
Hossana Twinomurizi	University of Johannesburg, South Africa
Oludayo O. Olugbara	Durban University of Technology, South Africa
Dustin Van der Haar	University of Johannesburg, South Africa
Cheikh Ahmadou Bamba Gueye	University Cheikh Anta Diop of Dakar, Senegal
Assane Gueye	Carnegie Mellon University Africa, Rwanda
Jean Marie Demebele	Gaston Berger University of Saint-Louis, Senegal
Youssou Faye	University Assane Seck of Ziguinchor, Senegal
Cheikh Sarr	University Iba Der Thiam of Thies, Senegal
Mamadou Bousso	University Iba Der Thiam of Thies, Senegal
Youssou Dieng	University Assane Seck of Ziguinchor, Senegal
Cheikh Talibouya Diop	Gaston Berger University of Saint-Louis, Senegal
Moussa Lo	Université virtuelle du Sénégal, Senegal
Dame Diongue	Gaston Berger University of Saint-Louis, Senegal
Ousmane Sall	Université virtuelle du Sénégal, Senegal
Ibrahima Niang	University Cheikh Anta Diop of Dakar, Senegal
Gervais Mendy	Ecole supérieure Polytechnique de Dakar, Senegal
Ado Adamou Abba Ari	University of Maroua, Cameroon
Hippolyte Michel T. Kenfack	University of Yaoundé I, Cameroon
Moussa Diallo	École supérieure Polytechnique de Dakar, Senegal
Mourad Gueroui	Université de Versailles Saint-Quentin en Yvelines, France
Jupiter Romuald N. Bakakeu	Alteryx, USA and University of Erlangen, Germany
Raphael Angulu	Masinde Muliro University of Science and Technology, Kenya
Glaucio H. S. Carvalho	Brock University, Canada
Andrea Visconti	Milan University, Italy
Herve Frackin	University of Le Havre, France
Lawrence Githuari	University of Nairobi, Kenya
Ismail Ateya	University of Strathmore, Kenya
Shibwabo Benard	University of Strathmore, Kenya
Mohammed-Sani Abdulai	Advanced Information Technology Institute, Ghana

Shadrack Maina Mambo	Kenyatta University, Kenya
Muhammad Fazal Ijaz	Sejong University, South Korea
Ayodele Adebisi	Covenant University, Nigeria
Muhammad Haroon Yousaf	University of Engineering & Technology Taxila, Pakistan
David Celestin Faye	Gaston Berger University of Saint-Louis, Senegal
Seydina Moussa Ndiaye	Université Virtuelle du Senegal, Senegal
Mary Ahuna	Technical University of Kenya, Kenya
Franklin Tchakounté	University of Ngaoundere, Cameroon
Mehrdad Tirandazian	Toronto Metropolitan University, Canada
Ignace Tchangou Toudjeu	Tshwane University of Technology, South Africa
Jimmi Talla Mbé	University of Dschang, Cameroon
Jules Waku	University of Yaoundé 1, Cameroon
Absalom Ezugwu	University of KwaZulu-Natal, South Africa
Justine Daramola	Cape Peninsula University of Technology, South Africa

Contents – Part I

Artificial Intelligence in Medicine

A Real-Time Monitoring System for Neurodegenerative Disease Progression Utilizing Generative Adversarial Networks	3
<i>Boriane Y. Tchaleu, Alain R. Ndjiongue, and Collins A. Leke</i>	
Detection and Classification of Prostate Cancer Using Lightweight MobileNet	15
<i>Olushola Olawuyi</i>	
A Comparative Analysis of Performance Metrics in Deep Learning Models for Brain Tumour Detection Using MRI Scans	34
<i>Larisha Moodley, Yuvika Singh, and Okuthe P. Kogeda</i>	
Prostate Cancer Detection and Classification Using Deep Learning Pretrained Model	54
<i>Olushola Olawuyi</i>	
Empowering Healthcare Queries: A Chatbot for Converting Text to SQL	72
<i>Nyashadzashe Tamuka and Khulumani Sibanda</i>	
HEALeB: Healthcare 4.0 Enabled by IoT, AI, and Blockchain	83
<i>Henriette Kyangwa, Landry Mbale, and Antoine Bagula</i>	

Smart Systems Enabling Technologies

Comparative Analysis of AI-Based Algorithms for the Detection of Tuberculosis Bacilli on Digitised Human Tissue	97
<i>Kapongo D. Lumamba, Threnesan Naidoo, Mpumelelo Msimnag, Gordon Wells, Adrie J. C. Steyn, and Mandlenkosi V. Gwetu</i>	
Classical vs Deep Learning Models: A Study on Maize Leaf Diseases Classification	121
<i>Chimango Nyasulu, Cecilia Tivir, and Jonathan Mwaura</i>	
Implementing a Novel Cybersecurity Algorithm for Smart Grid SCADA System	133
<i>Kholofelo H. Halefose, Topside E. Mathonsi, Tonderai Muchenje, and Deon Du Plessis</i>	

Improving Face Recognition from a Single Image 152
Justice Owusu Agyemang, Desmond Delali Atakpla, Awumee Gabriel Selorm, Thomas Boadu, Martin Kodua-Basoa, and Christabel Bemah Darko

Development of Automated Banana Sorter Using Computer Vision Techniques 166
Semakaleng Seretloa, Pius Owolawi, and Chunling Tu

Evaluating the Effectiveness of Current Countermeasures Against Man-In-The-Middle Attack on Domain Name System 177
Phandavhudzi Rotshidzwa, Moila Ramahlapane Lerato, and Velempini Mthulisi

Artificial Intelligence-Enabled Communication Systems

Predicting Coronal Hole Activity: Key to Mitigating Space Weather Impacts ... 193
Tarek Alsolame, B. D. S. Aritra, Enock Niyonkuru, Stephen Antogiovanni, and Chandranil Chakrabortii

Building Mapping for Implementing an Object-Oriented Model of Educational Infrastructure Usage at Jean Lorougnon Guédé (UJLoG) University in Central-Western Côte d’Ivoire 210
N’guessan Patrice Akoguhi, Kpangni Alex Jérémie Koua, Egomli Stanislas Assohoun, N’dri Pascal Kouame, and Koffi Fernand Kouame

Optical RIS in Indoor VLC Environments: What are the Suitable Materials 228
Frank Igboamalu, Alain Ndjiongue, Khmaies Ouahada, and Ling Cheng

Single- vs Multi-connectivity in Movable Access Point-Enabled Visible Light Communication Systems 240
Zohaib Saghir, Raphael Ahiaklo-Kuz, and Sylvester Aboagye

A Review on VLC-Based Indoor Positioning Technologies 252
Felipe Bueno

Evolving Trends in the Application of Artificial Intelligence to Support the Sustainable Development Goals: A Bibliometric Analysis 275
Kudakwashe Maguraushe, Patrick Ndayizigamiye, and Tebogo Bokaba

Author Index 293

Contents – Part II

Artificial Intelligence Theory and Methods

Scalable XAI: Towards Explainable Machine Learning Models in Distributed Systems	3
<i>Ibomoiye Domor Mienye, Theo G. Swart, and George Obaido</i>	
An Ensemble of Yolov5 Models in Real-Time Object Detection in Low Resource Settings	17
<i>Nkalubo Lenard Byenkya, Nakibuule Rose, and Okila Nixon</i>	
Hybridizing Deep Neural Networks and Machine Learning Models for Natural Scene Image Classification	36
<i>Clopas Kwenda, Mandlenkosi Victor Gwetu, and Jean Vincent Fonou Dombeu</i>	
Ensemble CNNs for Solar Flare Image Classification	55
<i>Mangaliso Mngomezulu and Mandlenkosi Gwetu</i>	
Natural Language Processing in Automatic Grading of Assessments in Higher Education: A Systematic Literature Review	70
<i>Lizzy Ofusori, Tebogo Bokaba, and Siyabonga Mhlongo</i>	
An Attention-Based Deep Learning Model for Term Extraction from Text Using BERT	87
<i>Tsitsi Zengeya, Jean Vincent Fonou Dombeu, and Mandlenkosi Gwetu</i>	

Artificial Intelligence and Smart Systems

Temporal Analysis of Social Concerns on African Social Media: Insights from Topics, Themes, and Sentiments	105
<i>Harriet Sibitenda, Awa Diattara, Assitan Traore, Elke Rundensteiner, and Cheikh Ba</i>	
Real-Time Object Detection Using an Ensemble of One Stage and Two Stage Object Detection Models with Dynamic Fine-Tuning Using Kullback-Leibler Divergence	125
<i>Nkalubo Lenard Byenkya, Nakibuule Rose, and Okila Nixon</i>	

Hybrid Approach for Image Processing Based on Convolutional Neural Network in Facial Recognition System	144
<i>Ibrahim Aoudou Pouamoun, Justin Moskolai Ngossaha, René Ndoundam, Francis Yongwa Dtissibe, Abdelhak Mourad Gueroui, and Ado Adamou Abba Ari</i>	
Network Anomaly Detection System Using an Attention-Based GNN	164
<i>Amir Mohammadi Bagha, Isaac Woungang, Issa Traore, Danda B. Rawat, Sudeep Tanwar, and Azin Hassanalizadeh</i>	
Truckchain: A Blockchain-Powered IoT Real-Time Tracking System for Fuel Supply Chain Management	177
<i>Cansa Kayembe Amaury, Kabemba Ntumbwa Jonathan, and Mpyana Mwamba Merlec</i>	
Food Safety 4.0: The Future of Food Safety Leveraging Industry 4.0 Technologies	194
<i>Eustace M. Dogo, Tebogo Bokaba, Hussaini A. Makun, Ahmed Aliyu, and Ponjul Bali Kparbong</i>	
Soil Drainage Classification Using Machine Learning Models: A Comparative Study	213
<i>Mouhamadou Lamine Diakhame, Diarra Bousso Dieng, Joshua Nougaret, and Jonathan Mwaura</i>	
A Comparative Analysis of FA1→3 and Lévy-Flight Based FA1→3 with Applications in Credit Card Fraud Using SMOTE Data Augmentation	227
<i>Trevor Vambe, Mandlenkosi Gwetu, and Serestina Viriri</i>	
Juro: A Retrieval-Augmented Generation AI Chatbot for Enhancing Legal Information Access in Resource-Constrained Settings	260
<i>Bernard Ngandu, Landry Mbale, and Antoine Bagula</i>	
Campus WiFi Demand Prediction: A Case Study on KNUST Campus	275
<i>Justice Owusu Agyemang, Manhyia Gideon Aboagye, Awumee Gabriel Selorm, Somuah Obed, Addai Clinton, and Henyo Chester Elikplim</i>	
Remote Sensing and Artificial Intelligence	
A Particle Swarm Optimization-Long-Short Term Memory (PSO-LSTM) Hybrid Model for Forecasting Global Horizontal Solar Radiation	291
<i>Blessing Olatunde Abisoye, Yanxia Sun, and Wang Zenghui</i>	

Modelling Road Networks with Node Structural Features and Graph Convolutional Networks	309
<i>Mohale Molefe</i>	
Leveraging Machine Learning and Climate Data for Enhanced Annual Crop Production Forecasts in Senegal	326
<i>Kpangni Alex Jérémie Koua, Fatou Cire Emilie Sané, Michael Warren, and Jonathan Mwaura</i>	
Road Obstacle Detection Using YOLO Algorithm Based on Attention Mechanism	347
<i>Bafokeng Lekola and Serestina Lekola</i>	
Machine Learning-Based System for Automated Crop Damage Detection and Classification in Insurance Claims	369
<i>Abdou Karim Kandji, Kpêchéhoué Merveille Santi Zinsou, Thomas Henahan, and Jonathan Mwaura</i>	
Author Index	387

Artificial Intelligence in Medicine



A Real-Time Monitoring System for Neurodegenerative Disease Progression Utilizing Generative Adversarial Networks

Boriane Y. Tchaleu¹✉, Alain R. Ndjiongue², and Collins A. Leke³

¹ Faculty of Engineering and Built Environment, University of Johannesburg, Corner Kingsway and University Road, Auckland Park 2006, South Africa

tboriane@gmail.com

² School of Electrical and Information Engineering, Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, South Africa

richard.ndjiongue@wits.ac.za

³ School of Accounting, University of Johannesburg, D-Ring 613, Auckland Park Kingsway Campus, Johannesburg, South Africa

collinsl@uj.ac.za

Abstract. Neurodegenerative diseases, such as Alzheimer's and Parkinson's, pose significant diagnostic and therapeutic challenges due to their complex and heterogeneous progression patterns. Current monitoring methods rely on periodic clinical assessments and neuroimaging, which may not capture the dynamic changes in disease pathology. This paper presents a novel approach for real-time monitoring of neurodegenerative disease progression using generative adversarial networks (GANs). Our method leverages longitudinal neuroimaging data to train a GAN-based model that generates synthetic images of brain structure and function, simulating disease progression. By comparing generated images with real-time neuroimaging data, we can detect subtle changes in disease pathology and track progression in real-time. We evaluate our approach on a large dataset of neuroimaging studies and demonstrate its potential for early detection, disease staging, and treatment efficacy monitoring. Our results show that GAN-based real-time monitoring can improve diagnostic accuracy and enable personalized treatment strategies for neurodegenerative diseases.

Keywords: Generative adversarial networks · Neurodegenerative diseases · Healthcare technology

1 Introduction

Neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease, and Huntington's disease, are a group of chronic and progressive conditions that affect millions of people worldwide [13]. These diseases are characterized by the

gradual degeneration of brain cells, leading to cognitive decline, motor dysfunction, and eventual loss of independence. Accurate disease progression modeling is crucial for early diagnosis, personalized treatment, and effective resource allocation. Despite significant advances in medical research, the diagnosis and treatment of neurodegenerative diseases remain a significant challenge. Current diagnostic methods rely heavily on clinical evaluations, imaging studies, and laboratory tests, which are often invasive, expensive, and provide only a snapshot of the disease state [10].

Real-time monitoring of neurodegenerative disease progression is crucial for optimal patient care and treatment planning. However, existing methods such as the approach in [1] relies on static magnetic resonance imaging (MRI) data and lacks personalization, similarly, the proposition of [11] enables to capture subtle changes in disease trajectory, while the technique in [7] lacks consideration of individual variability in disease progression, the work in [3] relies on limited training data and lacks robustness, and the methods proposed in [16] and [15] are still in their infancy and need more research and development.

Recently, GANs have emerged as a powerful tool in machine learning, capable of generating synthetic data that mimics real-world distributions [5]. Nevertheless, their potential for real-time monitoring of neurodegenerative diseases has not been fully explored.

Real-time monitoring of neurodegenerative diseases would enable early detection, accurate disease staging, and personalized treatment strategies. It would also allow for remote monitoring, reducing the burden on patients and health-care systems. Real-time monitoring would facilitate the identification of disease biomarkers, enabling the development of more effective treatments and improving patient outcomes. Moreover, it would provide a platform for monitoring treatment efficacy and disease progression in response to interventions.

In this paper, we propose a novel approach using GANs for real-time monitoring of neurodegenerative disease progression. Our approach integrates multi-modal data, including cognitive scores, imaging data, and wearable sensor outputs, to generate a continuous disease trajectory. We demonstrate the effectiveness of our approach in a retrospective study, showcasing accurate prediction of disease progression and detection of subtle changes in cognitive and motor functions.

The remainder of this paper is organized as follows: Sect. 2 provides a background on neurodegenerative diseases and existing monitoring methods. Section 3 introduces the proposed GAN architecture and our approach to real-time monitoring. Section 4 presents the results of our retrospective study. Section 5 discusses the implications of our findings, and concludes the paper.

2 Background on Neurodegenerative Diseases and Existing Monitoring Methods

It is estimated that approximately millions of people worldwide are affected by neurodegenerative diseases, which are characterized by progressive neuronal loss

and cognitive decline. The prevalence of these diseases is expected to triple by 2050, posing significant challenges to healthcare systems and economies globally [2]. The most common neurodegenerative diseases are: Alzheimer’s disease, Parkinson’s disease, Huntington’s disease, Amyotrophic lateral sclerosis, and Frontotemporal dementia. Neurodegenerative diseases share common pathophysiological mechanisms, including protein misfolding, oxidative stress, and neuroinflammation. Clinical features vary across diseases, but often include cognitive decline, motor dysfunction, and neuropsychiatric symptoms [3, 8].

Some existing neurodegenerative methods include those found in [4, 6, 9, 12]. For instance, the strategy described in the study by Fahimi *et al.* improves the robustness and performance of brain-computer interfaces (BCIs) by producing high-quality synthetic data to supplement scarce real data [4]. However, this approach has the risk of generating unrealistic or arte-factual data that may negatively impact BCI accuracy or lead to overfitting [4]. Helaly *et al.*’s deep learning approach enable accurate identification of subtle brain changes [6]. This approach offers promising results for early detection, potentially leading to timely interventions. Nevertheless, its effectiveness relies on large amounts of high-quality training data, which can be challenging to obtain, potentially impacting performance and generalizability [6]. The paper by Mukherjee *et al.* offers the advantage of generating high-quality synthetic brain images that can augment real data, improving the accuracy of disease diagnosis [9]. Nonetheless, the paper suffer from mode collapse, where GAN generates limited variations of synthetic images, reducing the diversity of the training dataset [9]. Shankar V.G. *et al.*’s approach uses GANs to reconstruct and enhance MRI slices, improving quality and resolution, and generating new training data [12]. This enables detection of subtle changes in brain structure and function, allowing for more accurate Alzheimer’s stage classification. However, the approach has limitations, including potential training instability, mode collapse, and challenging evaluation and clinical validation [12].

Current monitoring methods on neurodegenerative diseases rely on a combination of clinical assessments, neuroimaging, and biomarkers [14]. For example, some clinical assessments methods include cognitive tests (e.g. the mini-mental state examination (MMSE)), the Montreal cognitive assessment (MoCA), motor evaluations (e.g. the unified Parkinson’s disease rating scale (UPDRS)), and neuropsychiatric assessments (e.g. neuropsychiatric inventory). These clinical methods are subjective and prone to observe bias and patient variability. Another widely used methods is neuroimaging. Neuroimaging methods use structural MRI, functional MRI (fMRI), and positron emission tomography (PET) to provide valuable insights on the images for diseases modeling progression and diseases biomakers. However, they are often expensive, invasive, and require expert interpretation. Biomaker methods such as cerebrospinal fluid biomarkers (e.g. *amyloid* – β , tau) and blood biomarkers (e.g. inflammatory markers, neurofilament light chain) are invasive and may not accurately reflect disease progression [17].

Furthermore, general or commonly used methods such as electrophysiology (e.g. electroencephalography, electromyography) and genetic testing have limited applications in monitoring disease progression. These limitations highlight the need for more accurate, non-invasive, and cost-effective monitoring methods for neurodegenerative diseases. Hence the need of novel approaches, such as GANs for real-time monitoring of neurodegenerative disorders. This approach will offer numerous benefits, revolutionizing disease management and research. Furthermore, the approach enables early detection and diagnosis by identifying subtle changes in imaging data, cognitive tests, and biomarkers, allowing for timely interventions. Additionally, it enables real-time monitoring, while being robust to noise and missing data. A better understanding of disease progression can be obtained by modeling the complex relationships between biomarkers and progression. These methodologies have the potential to address these challenges and improve patient care.

3 Methodology

Each of the following stages, as shown in Fig. 1, requires a standardized protocols for image acquisition and processing. Furthermore, for security, the images are anonymized and store in a secure database.



Fig. 1. Methodology Workflow.

3.1 Data Collection

High-quality data collection is crucial for training and validating our GAN-based model for real-time monitoring of neurodegenerative diseases. We leverage a multi-modal dataset comprising neuroimaging data, clinical data, biomarker's data and genetic data. The neuroimaging data constitute of the longitudinal MRI and PET or computed tomography scans from the Alzheimer's disease neuroimaging initiative database. The datasets includes the MRI scans (T1-weighted, T2-weighted, and FLAIR),¹ functional MRI scans (resting-state fMRI

¹ The most common MRI sequences are T1-weighted and T2-weighted scans. Short repetition time and time-to-echo timings are used to create T1-weighted pictures. T1 characteristics of the tissue mostly determine the image's brightness and contrast. On the other hand, longer toxoplasmic encephalitis and time release periods result in T2-weighted pictures. Fluid-attenuated inversion recovery (FLAIR) is an MRI technique that shows areas of tissue T2 prolongation as bright while suppressing (darkening) cerebrospinal fluid (CSF) signal, thus clearly revealing lesions in proximity to CSF, such as cerebral cortical lesions.

and task-based fMRI), and PET scans (fluorodeoxyglucose and amyloid-PET). To obtain the clinical data, we collect demographic information, medical history, and cognitive assessments (MMSE, MoCA). Furthermore, we collect motor function tests/motor evaluations (UPDRS), and then collect other data such as neuropsychiatric assessments (neuropsychiatric inventory), medical history, medications, and demographics. For the biomarker's data, protein levels (e.g. amyloid-beta, tau), genetic markers (e.g. apolipoprotein, sudden cardiac arrest), and other biomarkers' data (e.g. inflammatory markers, oxidative stress markers) are collected. Genetic informations are collected from wearable sensors and mobile devices (e.g. heart rate, blood pressure), activity levels (e.g. step count, distance traveled), and sleep patterns (e.g. duration, quality). Effective data collection requires careful consideration of factors like data accuracy, completeness, consistency, privacy, and security to ensure reliable and informative data for subsequent processing and analysis.

3.2 Data Preprocessing

To ensure data quality and consistency, we perform image processing, feature extraction, data normalization, and data integration. Standard pipelines are utilized to preprocess the MRI and PET scans prior to image processing (e.g. skull stripping, normalization). Relevant features are extracted from clinical, biomarkers, and genetic data (e.g. cognitive scores, biomarker levels, genetic risk scores). All data are normalized to ensure consistent scales and distributions. Lastly, multi-modal data are integrated using a harmonization framework to ensure compatibility and consistency.

The preprocessed data are split into training (70%) for better machine learning rate, validation (15%), and testing sets (15%) to ensure robust model evaluation and generalizability. To increase the size and diversity of the training set, we apply data image and feature augmentation techniques. During image augmentation, we randomly flip, rotate, and crop MRI and PET scans. Moreover, Gaussian noise injection and feature masking are used for feature augmentation.

The training of a GAN-based model for real-time monitoring of neurodegenerative illnesses is made possible by this reliable pipeline for data collection and preprocessing, which capitalizes on the advantages of multi-modal data to enhance disease diagnosis, tracking, and treatment.

3.3 GAN Architecture

The GAN architecture is build using a deep learning framework (e.g. Tensorflow, Spyder) consisting of a generator and discriminator. The generator takes a random noise, z , as input and generates data that mimics real-world data. Examples of these data include synthetic medical imaging data (e.g. MRI, CT scans), synthetic clinical data (e.g. cognitive assessments, motor function tests), and synthetic biomarker's data (e.g. protein levels, genetic markers). The discriminator takes real-world data, medical data, clinical data, and biomarker data as input and predicts whether it is real or fake (i.e. generated by the generator).

Later, it provides feedback to the generator to improve the quality of the generated data. The generator and discriminator are trained simultaneously using a two-player game framework. The generator is trained to minimize the loss function (e.g. mean squared error, cross-entropy) between the generated data and the real-world data. The discriminator is trained to maximize the loss function (e.g. binary cross-entropy) between the real-world data and the generated data. An encoder network extracts features from real neuroimaging data to condition the generator. Given the following: $D(x_t)$ is the discriminator output for the real image, x_t , $G(z_t)$ is the created synthetic image at time t , z_t is the latent representation of x_t , and $D(G(z_t))$ is the discriminator output for the generated image $G(z_t)$. The adversarial loss or GAN loss function can be formulated as

$$L_{GAN} = \mathbb{E}_{x_t \sim p_{data}}[\log D(x_t)] + \mathbb{E}_{z_t \sim p_z}[\log(1 - D(G(z_t)))]. \quad (1)$$

The temporal consistency loss can be expressed as

$$L_{temp} = \mathbb{E}_{z_t \sim p_z, z_{t+1} \sim p_z}[\|G(z_t) - G(z_{t+1})\|_2^2], \quad (2)$$

where $G(z_{t+1})$ is the generated synthetic image at time $t+1$ and $\|a - b\|$ denotes the Euclidean distance from a to b . The reconstruction loss can be given by

$$L_{rec} = \mathbb{E}_{x_{t+1} \sim p_{data}}[\|x_{t+1} - \hat{x}_{t+1}\|_2^2], \quad (3)$$

where x_{t+1} is the real neuroimaging data at time $t+1$, and $\hat{x}_{t+1}$ is the reconstructed image at time $t+1$.

In total, adversarial loss, reconstruction loss, and temporal consistency loss are combined to train the GAN architecture. The overall loss function can be formulated as

$$L_{total} = L_{GAN} + \lambda_1 L_{rec} + \lambda_2 L_{temp}, \quad (4)$$

where λ_1 and λ_2 are hyperparameters that control the importance of each loss term, respectively. This total loss function complete the GAN architecture model for monitoring the diseases.

3.4 Real-Time Monitoring

During real-time monitoring, the trained GAN is used to generate synthetic data in real-time, based on incoming data from wearable sensors and mobile devices. The synthetic data is also used to predict diseases progression and detect disease biomarkers in real-time using a similarity metric (e.g. mean squared error, peak signal-to-noise ratio). As a specific example, Table 1 and Fig. 2, show Alzheimer's disease's progression visualization, representing the evolution of cognitive decline and neurological changes over time. These visualizations facilitate personalized medicine approaches, enabling clinicians to tailor treatment strategies to individual patient needs and monitor disease progression more effectively. Finally, new data is used to update the GAN model in real-time, allowing it to adjust to evolving disease patterns and enhance accuracy while facilitating early intervention and customized treatment. Given x_t as the real neuroimaging data at time

t , and $G(z_t)$ as the generated synthetic image at time t , the real-time monitoring is formulated as

$$\text{Monitor}[x_t, G(z_t)] = \begin{cases} 1, & \text{if } \|x_t - G(z_t)\|_2^2 > \epsilon, \\ 0, & \text{otherwise,} \end{cases} \quad (5)$$

where ϵ is a threshold value that controls the sensitivity of the monitoring system.

Table 1. Disease Progression Trajectories Visualization for example Alzheimer’s disease.

Mean Square Error Score (MMSE)	Time (Years)
30	0
27	1
23	2
19	3
15	4

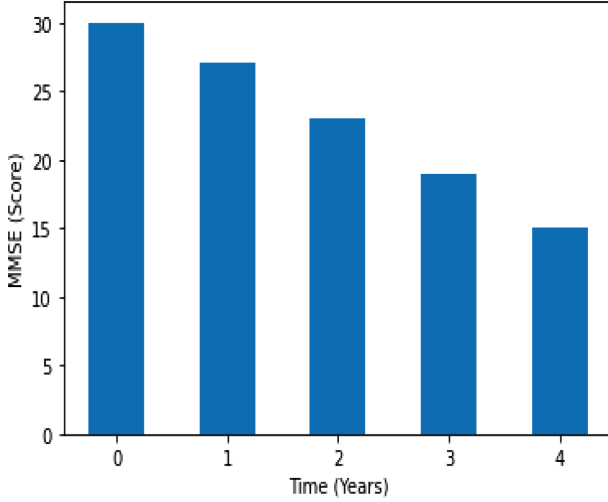


Fig. 2. Individual patient trajectories, illustrating his/her unique progression patterns.

4 Processing and Results

The percentage values of accuracy, sensitivity, specificity, and false positive rate are obtained by comparing the predictions made by the GAN model with the actual true labels of the data. The confusion matrix create a matrix to summarize the predictions versus actual true labels. This is expressed as

$$\text{Confusion matrix} = | \text{predicted positives} | \text{ predicted negatives} |, \quad (6)$$

where predicted positive means actual positive and predicted negative means actual negative. This equation is further expressed as

$$\mathbf{Actual\ positive} = | \text{true positive } (TP) | \text{false negatives } (FN) |, \quad (7)$$

and

$$\mathbf{Actual\ negative} = | \text{false positives } (FP) | \text{true negatives } (TN) |. \quad (8)$$

Following Eqs. 6, 7, and 8, the accuracy is calculated as the proportion of correct predictions $(TP + TN)$ out of total predictions $(TP + TN + FP + FN)$, and can be given by

$$\mathbf{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \times 100\%. \quad (9)$$

Sensitivity also called true positive rate is calculated as the proportion of true positives (TP) out of actual positive instances $(TP + FN)$, and can be given by

$$\mathbf{Sensitivity} = TP / (TP + FN) \times 100\%. \quad (10)$$

Specificity also called false positive rate is calculated as the proportion of true negatives (TN) out of actual negative instances $(TN + FP)$. Specificity can be written as

$$\mathbf{Specificity} = TN / (TN + FP) \times 100\%. \quad (11)$$

By applying these formulas to the confusion matrix, we can obtain the percentage values for accuracy, sensitivity, specificity, and false positive rate. For example, if the confusion matrix depicts an actual positive | 90 | 10 |, and actual negative | 5 | 95 |, as shown in Fig. 3, then:

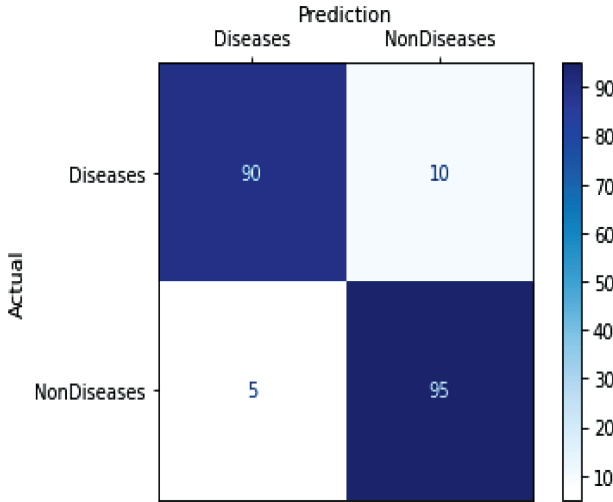


Fig. 3. Confusion matrix.

1. $Accuracy = (90 + 95) / (90 + 95 + 5 + 10) \times 100\% \approx 92.5\%$;
2. $Sensitivity = 90 / (90 + 10) \times 100\% \approx 90.0\%$;
3. $Specificity = 95 / (95 + 5) \times 100\% \approx 95.0\%$;

The above confusion matrices are also illustrated by Fig. 4.

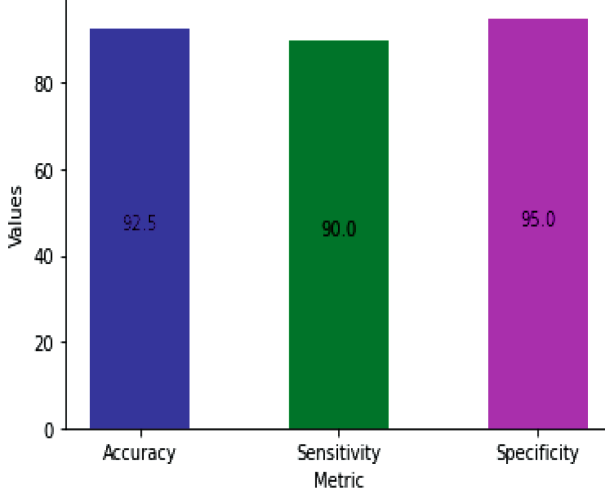


Fig. 4. Classification metrics.

Our GAN-based model detects subtle changes in disease pathology, enabling early detection of disease progression. Furthermore, the model accurately stages disease progression, allowing for personalized treatment strategies. Moreover, the model monitors treatment efficacy in real-time, enabling timely adjustments to treatment plans. Through the provision of individualized care and intervention. This strategy, help improve patient outcomes. To calculate the percentage of improvement using a large dataset of neuroimaging, we compared the performance metrics (*accuracy, sensitivity, specificity, and false positive rate*) of the GAN model with those of a baseline model or existing method. The formula to calculate the percentage of improvement is given by

$$\text{Percentage improvement} = ((\text{new value} - \text{old value}) / (\text{old value})) \times 100\%. \quad (12)$$

where new value is the performance metric value of the GAN model and old value is the performance metric value of the baseline model or existing method. For example, given that the baseline model has: an accuracy of 80%, a sensitivity of 70%, a specificity of 85%, and a false positive rate of 15%, and the GAN model has an accuracy of 95%, a sensitivity of 90%, a specificity of 95% and a false positive rate of 5%, the percentage of improvement can be calculated as

1. Accuracy improvement: $((95\% - 80\%) / 80\%) \times 100\% \approx 18.75\%$;
2. Sensitivity improvement: $((90\% - 70\%) / 70\%) \times 100\% \approx 28.57\%$;
3. Specificity improvement: $((95\% - 85\%) / 85\%) \times 100\% \approx 11.76\%$.

These quantitative calculations show that the GAN model has improved accuracy by 18.75% in detecting disease progression and has improved sensitivity by 28.57% in identifying minute changes in disease pathology. It has also improved specificity by 11.76% in detecting disease progression, compared to the baseline model as shown in Fig. 5. Additionally, a qualitative calculation of the model was done, which involves evaluating its performance using non-numerical measures, such as visual inspection, case studies, comparative analysis, error analysis, sensitivity analysis, and robustness analysis. These qualitative measures provide valuable insights into the model’s strengths, weaknesses, and limitations, complementing quantitative metrics like accuracy and sensitivity. Our method produces realistic-looking synthetic images that closely resemble actual neuroimaging data. Furthermore, our model recognizes minute changes in disease pathology and accurately models the development of the disease.

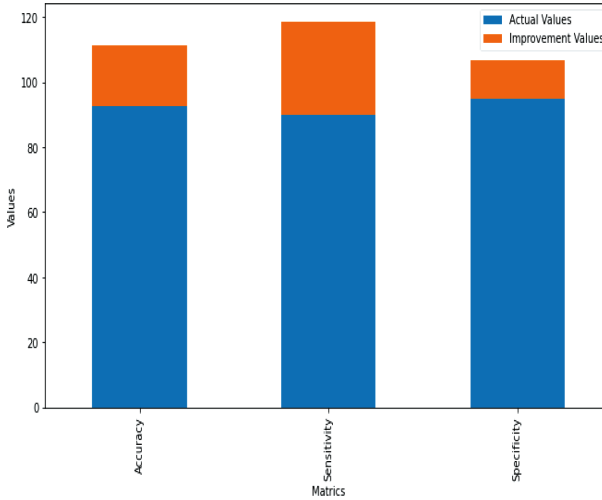


Fig. 5. Performance matrix.

Comparing our approach to the baseline model, we achieved 18.75% improvement in accuracy, 28.57% improvement in sensitivity, and reduced the false positive rate (specificity) by 11.76%. The approach demonstrates improved performance compared to existing methods, enabling early detection, accurate disease staging, and effective treatment efficacy monitoring.

5 Conclusion

In this paper, we introduced a unique GAN-based real-time monitoring technique for neurodegenerative diseases. A GAN-based model was trained to produce synthetic images of brain structure and function, simulating disease development

using longitudinal neuroimaging data. We identified minute changes in disease pathology and follow real-time progression by comparing produced images with real-time neuroimaging data. Our findings show that our strategy is successful in improving patient outcomes, precisely staging diseases, early disease identification, and treatment efficacy monitoring. Our method also offers the potential to transform the field of neurodegenerative disorders by facilitating quicker interventions, individualized treatment plans, and better patient outcomes. Our method has important ramifications for research, clinical practice, and enhancing patient outcomes and quality of life. Some limitations have been identified. For example, it is dependent on high-quality longitudinal neuroimaging data, computational requirements for training and inference, and need for further validation in larger and diverse cohorts. In conclusion, our GAN-based approach for real-time monitoring of neurodegenerative diseases has the potential to transform the of neuroscience by enabling early detection, accurate disease staging, and effective treatment efficacy monitoring. Our future work will expand to other neurodegenerative diseases (e.g., Parkinson's) and investigate transfer learning and domain adaptation techniques for improved generalizability.

References

1. Arunachalam, R., Sunitha, G., Shukla, S.K., Pandey, S.N., Urooj, S., Rawat, S.: A smart Alzheimer's patient monitoring system with IoT-assisted technology through enhanced deep learning approach. *Knowl. Inf. Syst.* **65**, 5561–5599 (2023)
2. Better, M.A.: Alzheimer's disease facts and figures. *Alzheimer's Dement.* **20**, 3708–3821 (2024)
3. Ciccocioppo, F., et al.: Neurodegenerative diseases as proteinopathies-driven immune disorders. *Neural Regen. Res.* **15**(5), 850–856 (2020)
4. Fahimi, F., Dosen, S., Ang, K.K., Mrachacz-Kersting, N., Guan, C.: Generative adversarial networks-based data augmentation for brain-computer interface. *IEEE Trans. Neural Netw. Learn. Syst.* **32**(9), 4039–4051 (2020)
5. Goodfellow, I., et al.: Generative adversarial networks. *Commun. ACM* **63**(11), 139–144 (2020)
6. Helaly, H.A., Badawy, M., Haikal, A.Y.: Deep learning approach for early detection of Alzheimer's disease. *Cogn. Comput.* **14**(5), 1711–1727 (2022)
7. Iqbal, M.S., et al.: Progress and trends in neurological disorders research based on deep learning. *Comput. Med. Imaging Graph.* 102–400 (2024)
8. Ma, M., Moulton, M.J., Lu, S., Bellen, H.J.: 'Flying' from rare to common neurodegenerative disease mechanisms. *Trends Genet.* **38**(9), 972–984 (2022)
9. Mukherjee, T., Sharma, S., Suganthi, K.: Alzheimer detection using deep convolutional GAN. In: *IEEE Madras Section Conference (MASCOS)*, Chennai, India, pp. 1–6 (2021)
10. Myszczyńska, M.A., et al.: Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nat. Rev. Neurol.* **16**(8), 440–456 (2020)
11. Noor, M., Zenia, N.Z., Kaiser, M.S., Mamun, S.A., Mahmud, M.: Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia. *Brain Inform.* **7**, 1–21 (2020)

12. Shankar, V.G., Sisodia, D.S.: Deep generative adversarial network-based MRI slices reconstruction and enhancement for Alzheimer's stages classification. In: *Advances in Deep Generative Models for Medical Artificial Intelligence*, pp. 65–82 (2023)
13. Sharma, S., Priya, T., Goel, N., Kashyap, D., Garg, V.K.: Overview of neurodegenerative disorders. In: *Data Analysis for Neurodegenerative Disorders*, pp. 3–12. Springer (2023)
14. Shusharina, N., et al.: Modern methods of diagnostics and treatment of neurodegenerative diseases and depression. *Diagnostics* **13**(3), 573 (2023)
15. Singh, K.U., More, A., Verma, R., Kumar Jain, D., Somasekar, J., Pandey, S.K.: Prediction of Alzheimer's disease progression using longitudinal brain MRI data and GANs. In: *International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI)*, Raipur, India, vol. 1, pp. 1–6 (2023)
16. Wang, Z., Gao, C., Glass, L.M., Sun, J.: Artificial intelligence for in silico clinical trials: a review. *arXiv* (2022)
17. Zhao, Y., Ma, B., Jiang, P., Zeng, D., Wang, X., Li, S.: Prediction of Alzheimer's disease progression with multi-information generative adversarial network. *IEEE J. Biomed. Health Inform.* **25**(3), 711–719 (2020)



Detection and Classification of Prostate Cancer Using Lightweight MobileNet

Olushola Olawuyi^(✉) 

School of Mathematics, Statistics and Computer Science University of
KwaZulu-Natal, Durban, South Africa
`221119659@stu.ukzn.ac.za`

Abstract. An early and precise diagnosis of prostate cancer (PCa) is essential to enhance treatment effectiveness. The detection of prostate cancer typically relies on human expertise and effort, which may sometimes lead to mistakes that could put a patient's life at risk. Therefore, the use of artificial intelligence (AI) using deep learning (DL) algorithms is proposed to assist experts in reducing diagnostic errors in the identification of prostate cancer. This research investigates how deep learning model, particularly MobileNet, can be used for automatic prostate cancer detection. MobileNet is known for its efficiency on mobile and embedded devices and offers a lightweight architecture that facilitates rapid image analysis with minimal computational resources. The proposed model was trained and tested using a magnetic resonance imaging (MRI) dataset containing both benign and malignant prostate tissue. The initial stages included data augmentation to enhance the diversity and quality of the training data. Performance metrics were used to evaluate the model during the training process, yielding the following outstanding results: accuracy (94%), precision (92%), recall (100%), and F1 score (96%) in identifying and categorizing prostate cancer. In summary, the research demonstrates how the deep learning model, specifically MobileNet, can enhance the effectiveness of detecting and classifying prostate cancer.

Keywords: Deep learning · Prostate cancer · MobileNet · Magnetic resonance imaging (MRI)

1 Introduction

Prostate cancer is highly prevalent among men worldwide, raising significant public health concerns. As a common cancer affecting men globally, it presents substantial challenges for early detection and accurate classification. Traditional diagnostic methods, such as histopathological examination of biopsy samples, rely heavily on pathologists' expertise and can be subject to variability and misinterpretation [11]. Recent advancements in AI, particularly in machine learning (ML) and deep learning (DL), have become powerful tools in healthcare, offering the potential to enhance the identification and assessment of prostate cancer [26]. These techniques utilize algorithms and computational methods to examine

and interpret large amounts of data. Various types of medical images, clinical records, and genetic data related to prostate cancer can be classified using these methods. The primary goal of using AI in the diagnosis of prostate cancer is to create predictive models that assist medical professionals in detecting potential cases, distinguishing between benign and malignant tumors, and selecting the best treatment options [13].

Deep learning models have offered promising solutions to enhance the accuracy and efficiency of prostate cancer diagnosis. This technique has been used to train large amounts of image data, compared to traditional machine learning models, leading to a more thorough understanding and practical application in real-life situations [31]. Among the various deep learning models, MobileNet and the Xception model stand out for their unique architectures and capabilities, making them suitable techniques for prostate cancer detection and classification. The availability of datasets and advancements in deep learning have enabled these systems to match healthcare professionals' performance on a variety of medical imaging tasks [12].

MobileNet is designed for efficiency and speed, particularly on mobile and embedded devices. By using depthwise separable convolutions, it reduces the number of parameters and computational load while still achieving high accuracy. This model is an excellent option for situations with limited computational resources where rapid and accurate analysis is required. The lightweight nature of MobileNet allows for deployment in point-of-care diagnostics and real-time monitoring systems, facilitating timely and accessible prostate cancer detection [4].

The examination of MRI images is essential for early disease detection and understanding. Deep learning models have demonstrated impressive capabilities in image classification tasks, and the MobileNet architecture has been extensively used across various fields [5,25,30]. On the other hand, Xception (Extreme Inception) represents a more sophisticated approach, building upon the Inception architecture. Xception utilizes depthwise separable convolutions, like MobileNet, but incorporates them into a more complex and deeper network structure [17]. Both MobileNet and Xception models have been extensively validated across various image recognition tasks, demonstrating their robustness and effectiveness [40].

Applying these models to prostate cancer detection involves training them on annotated medical image datasets, enabling them to learn and identify distinguishing features of cancerous and non-cancerous tissue. The use of data augmentation techniques as a pre-processing ensures that the models are well-equipped to handle variability in the data and generalize effectively to unseen samples.

The ability of the MobileNet model to be use for devices with low processing power, unlike deeper models such as ResNet or VGG, was a key motivation for choosing this model for this study. Therefore, we explore the use of MobileNet for identifying and categorizing prostate cancer (PCa).

2 Problem Statement

Men worldwide face prostate cancer challenges, which pose significant health risks if not detected and diagnosed early before complications occur. Recent research indicates that utilizing deep learning to identify prostate cancer and other health issues can alleviate the burden on healthcare professionals by minimizing errors in routine diagnoses, potentially reducing harm to patients. Our goal in this study is to develop a deep learning (DL) model that can detect prostate cancer in MRI images. The proposed model, MobileNet, will be tested for its accuracy in detecting prostate cancer and evaluated for its overall performance. This approach includes the use of MRI scans for prostate cancer detection, image processing techniques, and data analysis.

3 Objective

The objectives of this study are to develop a deep learning model (MobileNet) that can identify and categorize prostate cancer from MRI images and to evaluate its effectiveness in comparison with other state-of-the-art models in similar research.

4 Literature Review

The increased utilization of AI in the medical field has led to the development of a predictive computer-aided diagnosis tool for detecting PCa. Deep learning methods are particularly suited to MRI images because their process is non-invasive [7, 27, 28, 45]. The integration of multiparametric MRI and artificial intelligence allows CAD systems to deliver precise and rapid diagnoses of PCa, leading to enhanced patient care [41]. According to [6], a deep learning (DL) technique was suggested for the detection and boundary segmentation of the prostate using magnetic resonance images. The method achieved a Dice similarity coefficient of 0.89 with a mean absolute distance of 1.4 mm. Litjens et al. [23] created a computer-aided diagnosis (CAD) system that utilizes MRI images for screening. The results from the study showed 69% sensitivity and 83% specificity.

Similarly, Nitin et al. [37] proposed a deep learning method to segment and grade epithelial tissue, using a distinct approach that detects domain-agnostic characteristics. They examined whole slide images (WSI) from three groups of patients with PCa in their research. A total of 3741 core needle biopsies, collected from different sources, were used for the training procedure. The area under curve (AUC) and κ were computed to assess grade group comparison and evaluate the accuracy of core-level detection. An accuracy of 89.4% with a κ value of 0.92 was achieved on an internal test set of 425 CNB WSIs, whereas 85.3% accuracy with a κ value of 0.96 was observed on an external set of 1,201 images. When assessed on 1,303 WSIs from a third institution, the system achieved 83.1% accuracy with a κ value of 0.93.

Aldoj et al. [2] achieved an overall Dice similarity coefficient (DSC) of 0.92 by employing a DenseNet model, which resembles a U-Net structure, to automatically segment prostate zones from magnetic resonance imaging (MRI) scans. They enhanced segmentation by applying conditional random fields and cost-sensitive support vector machines (SVMs). A study by [3] introduced a computer-aided design system specifically created to locate prostate cancer on multi-spectral MRI images. The technique employed yielded a sensitivity of 88.9% and specificity of 92.2% in identifying cancer, as well as an average DSC value of 0.82.

A study by [32] developed a CAD method to identify and distinguish prostate cancer (PCa) from healthy tissue using mpMRI. They achieved an AUC of 0.85 when differentiating between cancerous and non-cancerous tissue. In a study by [16], the authors implemented a prostate segmentation technique that incorporated a customized convolutional neural network algorithm and used a training approach with statistical shape models, resulting in an average Dice similarity coefficient (DSC) of 0.895.

Tian et al. [42] introduced PSNet model for segmenting MRI prostate images. The proposed method achieved an average DSC of 0.94 with an average Hausdorff distance of 4.16 mm from the test dataset. The authors [24] suggested a technique for segmenting PCa using magnetic resonance images, which yielded mean DSC values of 0.815 for the prostate region and 0.666 for the cancerous regions.

The study of [1] utilized the VGG16 model with an ordinal classifier for Gleason scoring (GS) in grading PCa. Cao et al. [8] suggested a multiclass CNN and FocalNet for detecting PCa lesions and simultaneously estimating their aggressiveness using GS. They made use of attention models.

Hassan et al. [13] proposed an innovative classification algorithm that combines various deep learning techniques for the identification of PCa in ultrasound and MRI images. Customised layers were added to the pre-trained deep learning models, which were then trained on datasets. In the test set, the highest-performing model attained 97% accuracy on ultrasound (US) images and 80% accuracy on MRI images. The top performing model was selected to be used as a feature extractor to create a fusion model from the dataset in their future work.

In the research conducted by [35], the authors employed the YOLO versatile object detection algorithm to identify significant areas (for localization purposes) and evaluate those areas (for classification purposes). The algorithm was evaluated using a prostate cancer dataset generated by annotating 500 images of prostate tissue biopsies. The model was tested on two different datasets. The first dataset consisted of 50 images used for both training and testing, resulting in 97% accuracy for detection and classification. The second dataset included 137 prostate tissue images, and after training, the model achieved an accuracy of 89%.

Danyan et al. [22] used a deep learning model to enhance the diagnostic capability for prostate cancer by utilizing both multiparametric magnetic resonance imaging (mpMRI) and whole-mount histopathology data. A total of 739 par-

ticipants, comprising 466 with prostate cancer and 273 without, were enrolled between January 2017 and December 2019. The multiparametric magnetic resonance imaging data, specifically diffusion-weighted imaging, apparent diffusion coefficient sequences, and T2-weighted imaging, were divided into a training set (659) and a validation set (80). The deep learning (DL) model was later applied in real-world clinical settings (independent test dataset; $n = 200$), and its performance in detecting prostate cancer was compared between the DL model and radiologists of various expertise. The DL model diagnostic performance surpassed that of junior radiologists and demonstrated outstanding accuracy compared to both junior and senior radiologists' results.

Chin Wai et al. [9] evaluated the efficiency of three different deep learning (DL) convolutional neural network (CNN) algorithms in identifying prostate cancer. The Kaggle platform was used to simulate the three deep learning models: ResNet50, InceptionV3, and VGG16. These algorithms, which were used for various medical image diagnosis challenges, performed outstandingly in multiple competitions. A total of 620 image samples from the Cancer Imaging Archive (TCIA) dataset were utilized in their research. The performance of the three models was evaluated based on accuracy, F1 score, recall, and precision. The results showed that VGG16 achieved the highest accuracy at 95.56%, followed by ResNet50 at 86.67%, and InceptionV3 at 85.56%

Sun et al. [39] developed deep learning models to identify and localize clinically significant prostate cancer in multiparametric MRI images. Their models achieved an average AUC-ROC of 0.91 in identifying clinically significant PCa. To our knowledge, there is little literature focused on automated prostate cancer diagnosis, particularly using mpMRI images.

Ohad et al. [18] developed an advanced deep learning (DL) algorithm for diagnosing histopathology and grading Gleason scores in prostate biopsies. A urologic pathologist digitized 85 prostate core biopsy specimens obtained from 25 patients, annotating them for Gleason grades 3 & 4 or 5 prostate adenocarcinoma at $20\times$ magnification. The authors collected 14,803 image patches measuring 256 by 256 pixels, with a roughly equal distribution of malignancies. The deep residual convolutional neural network model was trained and tested for classifying patches at two levels: level one (coarse) as benign or malignant, and level two (fine) as benign, Gleason 3 & 4, and 5. The model's performance was evaluated using five-fold cross-validation. Randomization techniques were applied to compare the model's performance with random chance for hypothesis testing. The model achieved 91.5% accuracy ($p < 0.001$) in distinguishing between benign and malignant image patches at a general level, with 0.90 specificity, 0.93 sensitivity, and 0.95 average precision. Additionally, the model reached an 85.4% accuracy rate ($p < 0.001$) when categorizing image patches into benign, Gleason 3, Gleason 4, and Gleason 5, with 0.83 sensitivity, 0.94 specificity, and 0.83 average precision. Confusion mainly arose when differentiating between Gleason 3 & 4, and Gleason 4 & 5. Limitations include the limited sample size and the need for external validation.

Sanjay et al. [36] proposed a deep learning (DL) model to detect prostate cancer by utilizing the Gleason grading system on historical images. A 3D convolutional neural network (CNN) was employed to analyze and predict affected areas, supported by the Epithelial and Gleason grading network. The proposed model achieved state-of-the-art results, detecting both epithelial cells and Gleason scores simultaneously. The evaluation involved analyzing all MRI slices and volumes from the test fold, utilizing an endorectal coil to capture 3D MRI images of the prostate for segmentation of prostate cancer via a 3D CNN network. In experimental trials, this DL method achieved 85% specificity, 87% accuracy, and 89% sensitivity across different patient MRI images in the SPIE-AAPM-NCI Prostate dataset challenge.

Researchers in deep learning have focused on utilizing models such as U-Net, CNNs, and various customized models for the analysis of mpMRI images. Additionally, different computer-aided design (CAD) methods have been compared to deep learning architectures. Comparative studies consistently show that deep learning models outperform traditional CAD methods in medical image analysis, resulting in a significant advancement in the field [10, 29].

5 Methodology

5.1 CNN Concept

This section explores convolutional neural networks (CNNs), the MobileNet architecture and their application in classification tasks, providing a brief overview of both MobileNet and CNNs. It outlines the architecture and flowchart of MobileNet for prostate cancer classification. Object identification, categorization, and detection are primarily accomplished using deep learning-based CNN algorithms [15]. A typical CNN model consists of three layers: a convolutional layer, a pooling layer, and a fully connected layer [34]. The convolutional layer captures image characteristics and autonomously interprets image features. It uses multiple kernels and a specified stride to perform 3D convolution, generating feature maps directly from the input image [21]. Compared to other neural networks with similar features, CNNs have fewer parameters and connections, making it more efficient to train high-performing model. Different CNN algorithms vary in the construction of convolutional and pooling layers and the approach to network training [20]. In a convolutional layer, element-wise multiplication is performed for each input channel as the weight filter slides across the input, with the sum of those values saved sequentially in the output. This convolution process can be mathematically represented by Eq. (1), as shown in [43].

$$O_{x,y,n} = \sum_{c=1}^3 \sum_{j=1}^3 \sum_{i=1}^3 K_{i,j,c,n} \cdot F_{x+i-1,y+j-1,c} \quad (1)$$

In this context, O represents the output, K denotes the filter, F refers to the input, and i and j indicate the horizontal and vertical dimensions of the filter. Additionally, x and y represent the width and height of the input and output,

while c and n denote the input and output channels, respectively. The fundamental design of a CNN focuses on a portion of the input image rather than the entire image, as shown in Fig. 1, using a 3×3 grid to capture features. After the convolution process, max pooling is applied to reduce the size of the image stack. The max pooling layer, specified by its window size, scans the images with a filter, reducing the dimension of each feature map. Typically, CNN normalization is performed using a rectified linear unit (ReLU), which enhances the model's non-linear characteristics. The CNN also includes a fully connected layer, often referred to as the classifier, which is the final layer in terms of performance [33].

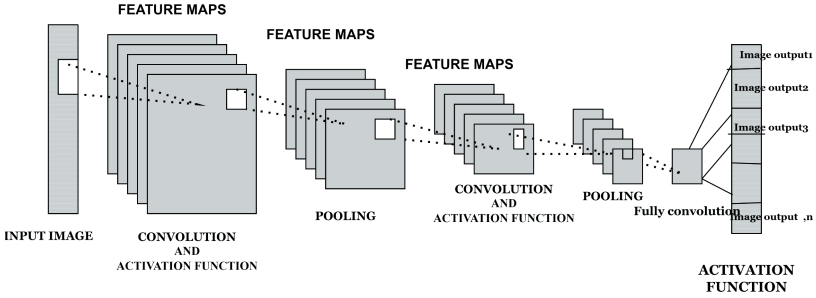


Fig. 1. Architecture of Convolutional Neural Network.

MobileNet Algorithm. Conventional CNN has been shown to perform very well in the area of image recognition; however, it requires a significant amount of computation for training and testing the algorithm. As a result, a high-performing graphics processing unit (GPU) and enough memory are necessary. MobileNet was developed to address this issue by modifying the convolution filter. Google developed MobileNet, which is optimized for small size and fast execution on mobile devices with limited resources, giving it a competitive advantage [21]. MobileNet utilizes depthwise separable convolution which differs from traditional convolution in its neural network design. Depthwise separable convolution breaks down a traditional convolution into two parts: filtering with depthwise convolution and summation processing with pointwise convolution. Depthwise convolution performs an individual filter operation for each channel within the input data before producing the final output. Pointwise convolution uses a 1×1 filter, which is then combined and output via depthwise convolution [14]. Equations (2) and (3) define the outcome of depthwise separable convolution when applied to RGB images using a 3×3 filter and a stride of 1 [43].

$$\hat{O} = \sum_{j=1}^3 \sum_{i=1}^3 K_{i,j,c} \cdot F_{x+i,y+j-1,c} \quad (2)$$

$$O = \sum_{i=1}^3 \hat{K}_{c,n} + \hat{O}_{x,y,c} \quad (3)$$

The output of the depthwise convolution is denoted as $\hat{O}$, with K representing the single filter used in the convolution and F representing the input, $\hat{O}$ is the result of the pointwise convolution; and $\hat{K}$ is the filter used in the 1×1 convolution. If the classic 3×3 convolution is replaced with a depthwise separable convolution, the computational workload decreases by approximately eight to nine times [14]. Figure 2 illustrates the overall structure of MobileNet, which includes 28 layers. The initial layer of the design features is 3×3 convolutions, which are used for basic filtering with one convolutional filter per input channel. The activation function known as the Rectified Linear Unit (ReLU) is faster and more accurate in computation compared to the sigmoid function. The Rectified Linear Unit function is applied in every convolutional layer [19].

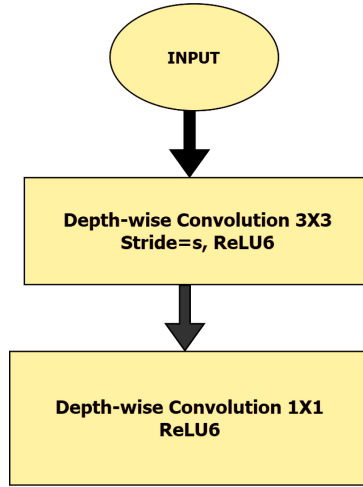


Fig. 2. MobileNet Architecture.

5.2 MobileNet Flow Chart for Cancer Classification

This research was conducted using the CNN MobileNet to identify prostate cancer from MRI images. The flowchart in Fig. 3 outlines the process of detection and classification of prostate cancer images with the MobileNet model. This model is faster than other image classification models as it uses a distinct filter for each response channel [44]. Two convolutional layers are constructed in the MobileNet algorithm to ensure accurate detection and classification of prostate cancer. The first layer is a depthwise convolution that applies one convolution on each input channel during filtering, while the second layer is a pointwise

convolution that combines the outputs of the depthwise convolution linearly. Adding batch normalization and a ReLU layer after each convolutional layer reduces computational cost, enabling faster classification compared to a standard CNN model by decreasing model size and complexity [38].

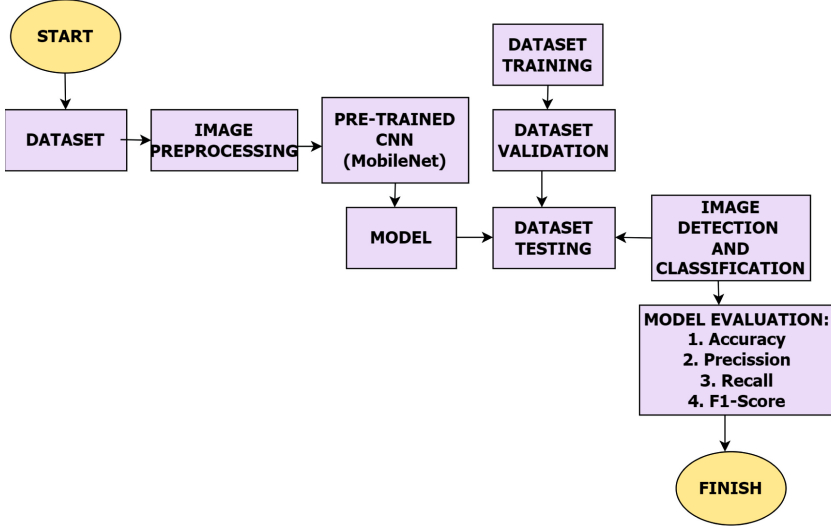


Fig. 3. Flow chart for the proposed model

5.3 Dataset

The study utilized MyDataset4 dataset from the Kaggle platform, which comprises MRI medical images sized 224×224 pixels. The dataset was divided into three categories-training, validation, and testing-for the detection and classification of prostate cancer. The dataset images were split into two classes: benign (non-cancerous) and malignant (cancerous). The dataset was divided into three sets: training (410 images), testing (180 images), and validation (30 images) for the two classes. Images were selected randomly during training. The training dataset includes 240 cancerous samples and 170 non-cancerous samples. In the testing phase, there are 60 non-cancerous and 120 cancerous samples, while in the validation phase, there are 10 non-cancerous and 20 cancerous samples. Figure 4 displays the dataset visualization, and Table 1 contains the dataset splitting information.

5.4 Pre-processing

Data augmentation was carried out as a data pre-processing step in this research. The augmentation technique ensures that the transformations of images are

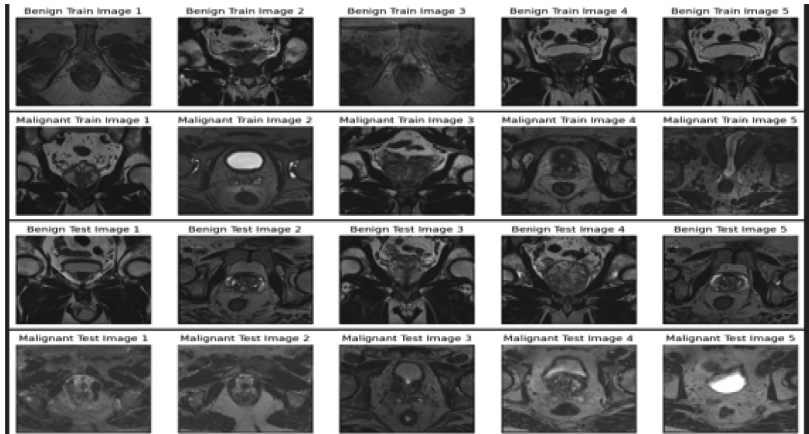


Fig. 4. Dataset Visualization

Table 1. Dataset Splitting into Training, Validation and Testing

Splitting	Number
Benign images in the training set	170
Malignant images in the training set	240
Benign images in the Validation set	10
Malignant images in Validation set	20
Benign images in the test set	60
Malignant images in the test set	120

being applied correctly and that the augmented images still represent realistic variations of the original image. Images from the dataset were shuffled randomly using a random function and augmentation was carried out before training the model. The original medical features were preserved in the enhanced images by flipping and rotating the initial images. Table 2 provides the parameters used for data augmentation along with their respective values.

Table 2. Augmentation Parameter

Parameter	Value
Rotation	10°
Zoom	0.1
Width and Height Shift range	0.2
Vertical and Horizontal flip	False
ZCA whitening	False
Sample and <i>featurewise_std_normalization</i>	False
Featurewise and Samplewise center	False

5.5 Algorithm1: Model Training of Prostrate Cancer Detection Using MRI Images

Input: MRI images input: $\rightarrow X = (x_1, x_2, \dots, x_N)$

Output: Images Detection $\rightarrow D = (D_1, D_2, \dots, D_N)$ and Classification Output $\rightarrow C = (C_1, C_2, \dots, C_N)$

Initialization: Hyperparameters and set E epochs, data loader $\rightarrow D = (D_{train}, D_{val}, D_{test})$ MobileNet model with random weights

for $K = 1 \rightarrow E$ do

1: Sample training data $I_k \in D_{train}$.

2. Obtain $I_{SK} \in \text{Mobilenet } (X_k)$.

3. Calculate the training loss value

4. Binary cross-entropy loss calculation.

5. Model performance evaluation on D_{val} for prostate cancer classification

end for

return model weights

5.6 Model Training, Validation, and Testing

The model was trained using MRI dataset divided into training, validation, and testing and was set to run over 100 epochs with a learning rate of 0.001, a batch size of 32, ReLU activation function for the input layer and Adam optimizer. After training, the model was evaluated on the testing dataset, Figs. 4 and 5 show the graphs illustrating the model accuracy and loss (Fig. 6 and Table 3).

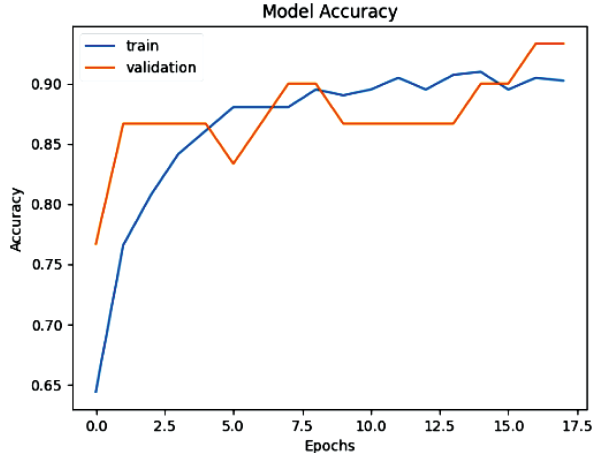


Fig. 5. Accuracy with Validation Accuracy

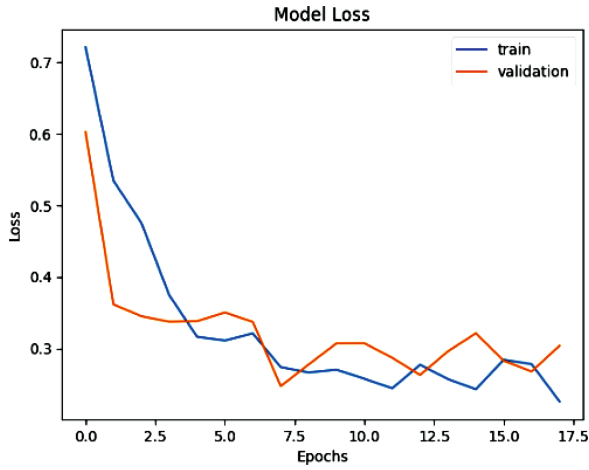


Fig. 6. Loss with Validation loss

Table 3. Numbers of trainable and non-trainable parameters.

Model	Total Parameters (Million)	Trainable Parameters	Non-trainable parameters (Million)
MobileNet	3,888,321	657,921	3,230,400

5.7 Model Prediction

After extensive training and testing, the model prediction are shown in Fig. 6. Prediction 0 indicates an image without cancer, while Prediction 1 indicates an image with cancer (Fig. 7).

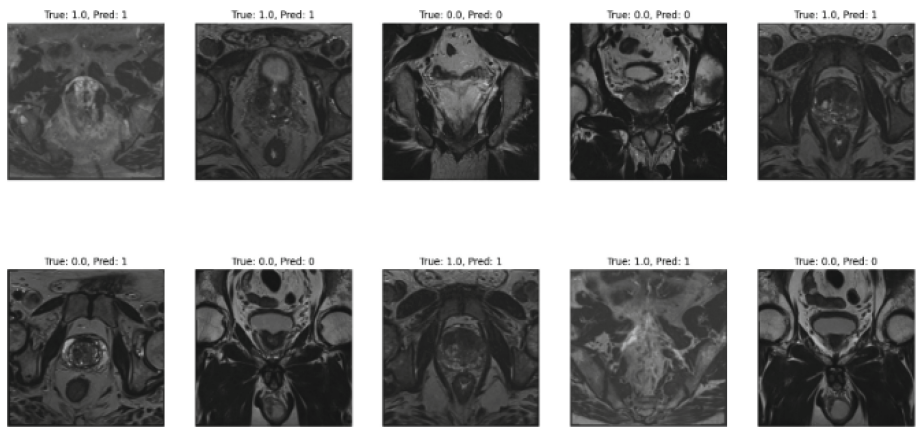


Fig. 7. Model prediction

The experiment was conducted on the Kaggle platform, which is designed for executing cloud-based deep learning projects using Jupyter notebooks. The platform is especially useful in the field of computer vision, allowing up to 20 GB of writable directory space and access to a $T4 \times 2$ GPU with dual processors, as shown in Table 4.

6 Hyper-Parameters

ReLU was used in the hidden layers, while Sigmoid was applied in the output layers with an initial learning rate of 0.001, dynamically adjusted to optimize the process. The Sigmoid function transforms outputs to a scale from 0 to 1, making it ideal for binary classification tasks. A batch size of 32 was chosen, as it is relatively standard and aligns well with the proposed model's efficient design. Adam stabilizes training, especially when used with a moderate learning rate (such as 0.001), by mitigating erratic gradient updates and reducing the risk of divergence. In MobileNet, binary cross-entropy and Adam work together effectively to optimize binary classification tasks (Table 5)

Table 4. Environment specification.

Platform	Programming language	Web application	Size of directory	GPU
Kaggle	Python 3.0	Jupyter notebook	20GB	T4 X 2

Table 5. Parameters settings.

Types of parameter	Value
Activation function for Input layer	Relu
Activation function for the output layer	sigmoid
Learning Rate	0.001
Batch size	32
Loss function	binary cross-entropy
Optimizer used	Adam
Number of categories	2

7 Evaluation

Evaluation criteria are crucial when creating and implementing DL models to assess their performance in diagnosis. Given TN as a true negative, TP as a true positive, FP as a false positive, and FN as a false negative, We evaluated performance of the proposed model in the following manner:

- Precision is calculated as the ratio of True Positive to the sum of True Positive and False Positive, represented by Eq. (5).
- Recall is calculated as the ratio of True Positive divided by the sum of True Positive and False Negatives according to Eq. (6).
- The formula for the F1-score is $2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$ as shown in Eq. (7).

$$ACCURACY = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

$$PRECISION = \frac{TP}{TP + FP} \quad (5)$$

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

$$F1SCORE = \frac{2 * Precision * recall}{Precision + Recall} \quad (7)$$

8 Results and Discussion

The proposed MobileNet model was evaluated using four metrics: accuracy (94%), precision (92%), recall (100%), and F1-score (96%). It successfully provided accurate predictions for benign and malignant tumors. The loss value was also used for model evaluation during the training process, with binary cross-entropy as the adopted loss function. At epoch 1 of 100, the loss value was initially high at 1.0, indicating significant model error. However, as training progressed to epoch 40, the loss value decreased to 0.20. At this point, the model's accuracy stabilized, with no further improvement in performance (Tables 6 and 7).

Table 6. Model result.

Accuracy	Precision	Recall	F1-score
94%	92%	100%	96%

9 Conclusion

This study proposed a technique for identifying and classifying prostate cancer by utilizing the state-of-the-art MobileNet model. The lightweight, intelligent model was able to distinguish between cancerous and non-cancerous lesions using MRI dataset. This model is suitable for low-configuration computers, making it advantageous for units with limited computer resources. To detect and classify prostate cancer, the power of the pre-trained MobileNet model was analyzed.

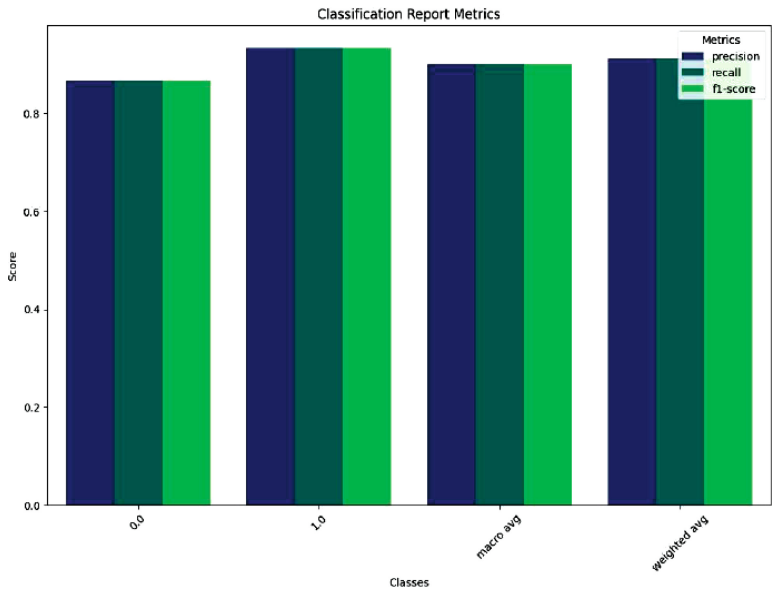


Fig. 8. Classification representation

Table 7. Comparison of deep learning models’ accuracy

Ref	Model	Dataset Attribute	Purpose	Accuracy(%)
[13]	Deep learning	MRI images	Prostate cancer classification	80
[37]	Deep learning	Whole slide images (WSI)	Prostate cancer diagnosis	89.4, 85.3, 83.1
[35]	Deep learning (Yolo algorithm)	Prostate tissue images	Prostate cancer grading and diagnosis	89
[22]	Deep learning	Multiparametric Magnetic Resonance	Prostate cancer diagnosis	70
[9]	Deep learning (ResNet50 and InceptionV3)	MRI images	Prediction Cancer.	86.67 & 85.56.
[18]	Deep learning	Histopathological image	Histopathologic Diagnosis and Gleason grading of prostate cancer	85.4
[36]	Deep learning framework	MRI	Detection of prostate cancer	87
Proposed Model	Deep learning (MobileNet)	MRI images	Prostate cancer detection	94

The prediction was validated using a publicly available MRI dataset, and the accuracy of the model was 94%. The model’s performance will enable analysts to generate ideas for developing more effective CNN-based models for the initial treatment of prostate cancer. Unfortunately, limited data and privacy issues are significant challenges in this field due to confidentiality concerns. There has

been a lack of available datasets for prostate cancer. Nevertheless, to ensure the validity of the data, multiple experts need to annotate each image. Inadequate datasets impede training and can result in model overfitting. Our goal is to further optimize the proposed model to achieve improved results. This will help specialists rapidly and effectively diagnose prostate abnormalities.

Furthermore, in the future, this lightweight model could potentially be integrated into mobile devices due to its small size and limited parameters, enabling mobile users to perform self-screening.

References

1. Abraham, B., Nair, M.S.: Automated grading of prostate cancer using convolutional neural network and ordinal class classifier. *Inf. Med. Unlocked* **17**, 100256 (2019). <https://doi.org/10.1016/j.imu.2019.100256>
2. Aldoj, N., Biavati, F., Michallek, F., Stober, S., Dewey, M.: Automatic prostate and prostate zones segmentation of magnetic resonance images using DenseNet-like U-net. *Sci. Rep.* **10**(1), 14315 (2020). <https://doi.org/10.1038/s41598-020-71080-0>
3. Artan, Y., et al.: Prostate cancer localization with multispectral MRI using cost-sensitive support vector machines and conditional random fields. *IEEE Trans. Image Process.* **19**(9), 2444–2455 (2010). <https://doi.org/10.1109/TIP.2010.2048612>
4. Athiramol, S.: Enhanced MobileNet architecture with residual blocks for improved CT image classification. In: 2024 1st International Conference on Trends in Engineering Systems and Technologies (ICTEST), pp. 01–05. IEEE (2024)
5. Bhatia, S., Sinha, Y., Goel, L.: Lung cancer detection: a deep learning approach. In: Bansal, J.C., Das, K.N., Nagar, A., Deep, K., Ojha, A.K. (eds.) *Soft Computing for Problem Solving. AISC*, vol. 817, pp. 699–705. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-1595-4_55
6. Brosch, T., Peters, J., Groth, A., Stehle, T., Weese, J.: Deep learning-based boundary detection for model-based segmentation with application to MR prostate segmentation. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) *MICCAI 2018. LNCS*, vol. 11073, pp. 515–522. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00937-3_59
7. Cao, R., et al.: Joint prostate cancer detection and Gleason score prediction in mp-MRI via FocalNet. *IEEE Trans. Med. Imag.* **38**(11), 2496–2506 (2019)
8. Cao, R., et al.: Joint prostate cancer detection and Gleason score prediction in mp-MRI via FocalNet. *IEEE Trans. Med. Imag.* **38**(11), 2496–2506 (2019). <https://doi.org/10.1109/TMI.2019.2901928>
9. Chun, C.W., Yousir, N.T., Abdulameer, S.M., Mostafa, S.A., Hezam, A.A.: Deep learning approach for predicting prostate cancer from MRI images. *J. Soft Comput. Data Min.* **3**(2), 1–9 (2022)
10. Das, S., Majumder, S.: Lung cancer detection using deep learning network: a comparative analysis. In: 2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), pp. 30–35 (2020). <https://doi.org/10.1109/ICRCICN50933.2020.9296197>
11. Gorelick, L., et al.: Prostate histopathology: learning tissue component histograms for cancer detection and classification. *IEEE Trans. Med. Imag.* **32**(10), 1804–1818 (2013)

12. Gulshan, V., et al.: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **316**(22), 2402–2410 (2016)
13. Hassan, M.R., et al.: Prostate cancer classification from ultrasound and MRI images using deep learning based explainable artificial intelligence. *Futur. Gener. Comput. Syst.* **127**, 462–472 (2022)
14. Howard, A.G., et al.: MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv preprint [arXiv:1704.04861](https://arxiv.org/abs/1704.04861) (2017)
15. Hwang, S., Park, J., Won, J., Kwon, Y., Kim, Y.: Object detection for cargo unloading system based on Fuzzy C means. *Comput. Mater. Continua* **71**(2), 4167–4181 (2022). <https://doi.org/10.32604/cmc.2022.023295> <https://doi.org/10.32604/cmc.2022.023295>
16. Karimi, D., Samei, G., Kesch, C., Nir, G., Salcudean, S.E.: Prostate segmentation in MRI using a convolutional neural network architecture and training strategy based on statistical shape models. *Int. J. Comput. Assist. Radiol. Surg.* **13**(8), 1211–1219 (2018). <https://doi.org/10.1007/s11548-018-1785-8>
17. Khan, Z., Yahya, N., Alsailh, K., Meriaudeau, F.: Segmentation of prostate in MRI images using depth separable convolution operations. In: Singh, M., Kang, D.-K., Lee, J.-H., Tiwary, U.S., Singh, D., Chung, W.-Y. (eds.) *IHCI 2020. LNCS*, vol. 12615, pp. 132–141. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-68449-5_14
18. Kott, O., et al.: Development of a deep learning algorithm for the histopathologic diagnosis and Gleason grading of prostate cancer biopsies: a pilot study. *Euro. Urol. Focus* **7**(2), 347–351 (2021). <https://doi.org/10.1016/j.euf.2019.11.003>
19. Krizhevsky, A., Sutskever, I.H.G.: ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25** (2012)
20. Lee, S.H.: Deep learning based face mask recognition for access control. *J. Korea Acad. Ind. Coop. Soc. Korea Sci.* **21**(8), 395–400 (2020)
21. Lee, H., Grosse, R., Ranganath, R., Ng, A.Y.: Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 609–616. ICML '09, Association for Computing Machinery, New York, NY, USA (2009). <https://doi.org/10.1145/1553374.1553453>
22. Li, D., et al.: Deep learning in prostate cancer diagnosis using multiparametric magnetic resonance imaging with whole-mount histopathology referenced delineations. *Front. Med.* **8**, 810995 (2022)
23. Litjens, G., Debats, O., Barentsz, J., Karssemeijer, N., Huisman, H.: Computer-aided detection of prostate cancer in MRI. *IEEE Trans. Med. Imag.* **33**(5), 1083–1092 (2014). <https://doi.org/10.1109/TMI.2014.2303821>
24. Liu, X., Langer, D.L., Haider, M.A., Yang, Y., Wernick, M.N., Yetik, I.S.: Prostate cancer segmentation with simultaneous estimation of Markov random field parameters and class. *IEEE Trans. Med. Imag.* **28**(6), 906–915 (2009). <https://doi.org/10.1109/TMI.2009.2012888>
25. Mathews, R.P., Panicker, M.R.: Towards fast region adaptive ultrasound beam-former for plane wave imaging using convolutional neural networks. In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 2910–2913. IEEE (2021)
26. Matte, S., et al.: Prostate cancer detection using deep learning and traditional techniques. *IEEE Access* **9**, 27085–27100 (2024)

27. Mehta, P., Antonelli, M., Ahmed, H.U., Emberton, M., Punwani, S., Ourselin, S.: Computer-aided diagnosis of prostate cancer using multiparametric MRI and clinical features: a patient-level classification framework. *Med. Image Anal.* **73**, 102153 (2021)
28. Mehta, P., et al.: AutoProstate: towards automated reporting of prostate MRI for prostate cancer assessment using deep learning. *Cancers* **13**(23), 6138 (2021). <https://doi.org/10.3390/cancers13236138>
29. Ming, J.T.C., Noor, N.M., Rijal, O.M., Kassim, R.M., Yunus, A.: lung disease classification using different deep learning architectures and principal component analysis. In: 2018 2nd International Conference on BioSignal Analysis, Processing and Systems (ICBAPS), pp. 187–190 (2018). <https://doi.org/10.1109/ICBAPS.2018.8527385>
30. Patil, V.C., Dhotre, S.R.: Lung cancer detection from images of computer tomography scan. *Int. J. Adv. Res. Comput. Commun. Eng.* **5**(7) (2016)
31. Peng, H., Xu, H., Shen, G., Liu, H., Guan, X., Li, M.: A lightweight crop pest classification method based on improved MobileNet-V2 model. *Agronomy* **14**(6), 1334 (2024)
32. Peng, Y., et al.: Quantitative analysis of multiparametric prostate MR images: differentiation between prostate cancer and normal tissue and correlation with Gleason score—a computer-aided diagnosis development study. *Radiology* **267**(3), 787–796 (2013). <https://doi.org/10.1148/radiol.13121454>
33. Praveen Gujjar, J., Prasanna Kumar, H.R., Chiplunkar, N.N.: Image classification and prediction using transfer learning in colab notebook. *Glob. Transitions Proc.* **2**, 382–385 (2021). <https://doi.org/10.1016/j.gltp.2021.08.068>
34. Qian, Y., Dong, J., Wang, W., Tan, T.: Deep learning for steganalysis via convolutional neural networks. In: *Media Watermarking, Security, and Forensics 2015*, vol. 9409, pp. 171–180. SPIE (2015)
35. Salman, M.E., Cakirsoy Cakar, G., Azimjonov, J., Kösem, M., Cedimoglu, I.H.: Automated prostate cancer grading and diagnosis system using deep learning-based Yolo object detection algorithm. *Expert Syst. Appl.* **201**, 117148 (2022). <https://doi.org/10.1016/j.eswa.2022.117148>
36. Singh, S.K., et al.: A novel deep learning-based technique for detecting prostate cancer in MRI images. *Multimedia Tools Appl.* **83**(5), 14173–14187 (2023). <https://doi.org/10.1007/s11042-023-15793-0>
37. Singhal, N., et al.: A deep learning system for prostate cancer diagnosis and grading in whole slide images of core needle biopsies. *Sci. Rep.* **12**(1), 3383 (2022). <https://doi.org/10.1038/s41598-022-07217-0>
38. So, M.H., Han, C.S., Kim, H.Y.: Defect classification algorithm of fruits using modified MobileNet. *J. Korean Inst. Inf. Technol.* **18**(7), 81–89 (2020)
39. Sun, Z., et al.: Deep-learning models for detection and localization of visible clinically significant prostate cancer on multi-parametric MRI. *J. Mag. Reson. Imag.* **58**(4), 1067–1081 (2023). <https://doi.org/10.1002/jmri.28608>
40. Sutaji, D., Yıldız, O.: LEMOXINET: Lite ensemble MobileNetV2 and Xception models to predict plant disease. *Eco. Inform.* **70**, 101698 (2022)
41. Tabatabaei, Z., Colomer, A., Engan, K., Oliver, J., Naranjo, V.: Residual block convolutional auto encoder in content-based medical image retrieval. In: 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), pp. 1–5 (2022). <https://doi.org/10.1109/IVMSP54334.2022.9816325>
42. Tian, Z., Liu, L., Zhang, Z., Fei, B.: PSNet: prostate segmentation on MRI based on a convolutional neural network. *J. Med. Imag.* **5**(2), 021208–021208 (2018)

43. Wang, H., Lu, F., Tong, X., Gao, X., Wang, L., Liao, Z.: A model for detecting safety hazards in key electrical sites based on hybrid attention mechanisms and lightweight Mobilenet. *Energy Rep.* **7**, 716–724 (2021)
44. Wu, T.Y., et al.: Identifying Fagaceae and Lauraceae species using leaf images and convolutional neural networks. *Ecol. Inf.* **68**, 101513 (2022)
45. Zhong, X., et al.: Deep transfer learning-based prostate cancer classification using 3 Tesla multi-parametric MRI. *Abdom. Radiol.* **44**, 2030–2039 (2019)



A Comparative Analysis of Performance Metrics in Deep Learning Models for Brain Tumour Detection Using MRI Scans

Larisha Moodley, Yuvika Singh, and Okuthe P. Kogeda^(✉) 

School of Mathematics, Statistics and Computer Science, College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Westville Campus, Durban 3209, Republic of South Africa

220036955@stu.ukzn.ac.za, {SinghY1,KogedaO}@ukzn.ac.za

Abstract. Brain cancer is one of the deadliest diseases affecting the brain. Medical imaging techniques are commonly employed for the identification of brain tumours, and Deep Learning is rapidly emerging as a powerful tool in this detection process. This study evaluates the efficiency of Deep Learning Convolutional Neural Networks (CNNs)—including ResNet50, EfficientNetB0, VGG16, and InceptionV3—against traditional Machine Learning models such as Logistic Regression, Support Vector Machine (SVM), and Random Forest for detecting brain tumours from Magnetic Resonance Imaging (MRI) scans. Automated detection techniques offer significant improvements in patient outcomes, particularly given the increasing incidence of brain tumours and the urgent need for swift, accurate diagnoses to save lives. To address this challenge, the study utilised advanced CNN architectures to analyse the “Br35H: Brain Tumour Detection 2020” dataset of brain MRI images with annotations. The performance of the models was assessed based on accuracy, precision, recall (sensitivity), specificity, F1-score, and ROC–AUC metrics. The results demonstrated that Deep Learning models achieved significantly higher accuracy compared to traditional Machine Learning methods, with ResNet50 yielding the best results: 99.33% accuracy, 99.11% precision, and 99.55% recall.

Keywords: Brain Tumour · Convolutional Neural Network (CNN) · Deep Learning (DL) · EfficientNetb0 · InceptionV3 · Logistic Regression · Machine Learning (ML) · Random Forest · ResNet50 · Support Vector Machine (SVM) · VGG16

1 Introduction

The brain is an organ composed of trillions of cells that control all functions of the body. Tumours can arise from uncontrolled cell division, leading to the formation of abnormal masses within the brain or surrounding tissues. Brain tumours are prevalent in both adults and children and exhibit the highest morbidity and mortality rates among cancers [1–4].

Their impact on patient survivability and quality of life presents a significant global health concern. Tumours are classified into two main categories: malignant and benign. Benign (non-cancerous) tumours, characterised by slow growth and a low probability of recurrence with appropriate treatment, contrast sharply with malignant (cancerous) tumours, which spread rapidly and can invade surrounding tissues in a process called metastasis [5].

Early detection and intervention are crucial for effectively treating and managing brain tumours. Magnetic Resonance Imaging (MRI) is one of the foremost diagnostic methods, providing high soft tissue contrast and detailed anatomical images essential for identifying and assessing brain malignancies. However, radiologists' manual interpretation of these scans is often tedious and prone to human error [6]. To address this issue, implementing automated image processing systems has garnered increased attention [7–9]. Recent advancements in Deep Learning have shown promising results in medical image processing, particularly in tumour identification. One notable technique is Convolutional Neural Networks (CNNs), which were first successfully applied by Yann LeCun in 1998 to classify handwritten digits [10]. The ImageNet initiative, launched in 2010, has further advanced CNN research through the Large Scale Visual Recognition Challenge (ILSVRC), which annually tests algorithms on the provided datasets [11].

CNN architectures have proven capable of detecting various patterns and features in MRI scans, enabling highly accurate and efficient automated detection. These networks mimic the multi-layered structure of the brain's visual cortex, employing convolutional layers to automatically identify hierarchical image features such as edges, textures, and shapes. Models like VGG16, Inception, and ResNet have gained popularity in medical imaging tasks due to their integrated feature extraction and classification capabilities within an end-to-end learning framework [12]. The intersection of medical imaging and artificial intelligence has revolutionised diagnostic medicine, facilitating the identification of various clinical conditions, including brain tumours. The World Health Organization (WHO) classifies brain tumours based on health behaviour, categorising low-grade (benign) tumours as grades 1 and 2 and high-grade (malignant) tumours as grades 3 and 4 [13]. It is estimated that approximately 250,000 individuals are diagnosed with brain tumours annually, with only 2% classified as malignant [14].

Detecting brain tumours earlier and more accurately can significantly enhance patient quality of life through MRI scans. However, challenges remain in achieving reliable and high-efficiency detection performance. Accurate identification of brain tumours is essential for improving patient survival rates and guiding treatment options. Automated identification systems face hurdles due to the variability and complexity of tumour characteristics [15]. Therefore, developing sophisticated CNN architectures and optimisation techniques tailored to medical imaging data is necessary. Despite the vast potential of Deep Learning approaches, challenges persist in optimising CNN architectures for brain tumour diagnosis. Issues such as the lack of annotated datasets, the risk of model overfitting, and difficulties in interpreting findings complicate developing and implementing robust detection systems. Addressing these challenges requires a thorough understanding of brain tumours' biology and imaging characteristics, along with innovative approaches for model development and evaluation.

In addition to Deep Learning, traditional Machine Learning algorithms have also been employed to classify brain tumours using features extracted from MRI scans [16]. These include Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbour (KNN). While effective, these models often rely on handcrafted features, which can limit their flexibility. Hybrid methods that integrate Deep Learning's automatic feature extraction with traditional Machine Learning's established classification techniques have succeeded in improving accuracy and generalizability in brain tumour classification [17]. This dual approach mitigates challenges associated with exclusive reliance on Deep Learning models, enhancing interpretability and reducing overfitting.

This study used an annotated brain MRI dataset to evaluate the performance of various CNN architectures and traditional Machine Learning models in terms of their ability to generalise unseen data. The aim was to enhance patient prognosis and quality of life through comprehensive experimental and comparative analysis by employing Computer-aided Detection (CAD). The study further sought to identify the limitations and strengths of each model through relevant experiments and comparative analyses. The main contributions of this study include:

- A comparative study of advanced Deep Learning models (ResNet50, EfficientNetB0, VGG16, and InceptionV3) and traditional Machine Learning methods (Logistic Regression, SVM, Random Forest) for brain tumour detection using MRI scans.
- An evaluation of each approach in relation to precision, accuracy, recall (sensitivity), specificity, F1- score, and ROC–AUC metrics to determine overall performance.
- Findings indicating that Deep Learning models, particularly ResNet50, demonstrate superior performance compared to traditional methods, achieving an accuracy of 99.33% and precision of 99.11%. These results illustrate the potential of CNNs for consistent and accurate diagnoses in medical applications.
- An examination of model generalisation and efficiency showed that Deep Learning models provide better generalised performance, improved computational efficiency, minimal signs of overfitting, and faster convergence than conventional methods.
- Identification of practical challenges such as overfitting, limited GPU utilisation during training, and the complexities of optimising hyperparameters. These insights offer direction for researchers to address these challenges and enhance automated detection systems.

The remainder of this paper is structured as follows: In Sect. 2, literature review is provided. In Sect. 3, details of the dataset, data preprocessing techniques, and the CNN and traditional machine learning models employed in the study are discussed. In Sect. 4, evaluation criteria, the proposed experimental design, and the results and performance analysis of each model, including the limitations of the study and implications for future research. Finally, in Sect. 5, a summary of the main accomplishments of the research and offers suggestions for further research.

2 Literature Review

In the field of medical image processing, it is known that CNNs have been used to enhance the detection of brain tumours by automating brain tumour diagnosis using magnetic resonance imaging (MRI) images. MRI scans are a popular method for imaging brain regions due to its ability in determining different types of tissue and being the most versatile [18, 19]. Many research works have been conducted with an aim of properly understanding and developing CNN architectures like VGG16, Inception, ResNet, DenseNet, MobileNet, EfficientNet, and AlexNet, with some being more successful than others.

A proposed model for assessing medical brain images and identifying intricate and hidden aspects for the diagnosis of brain tumour disease makes use of deep Convolutional Neural Networks [20]. The accuracy of the suggested model increased to 94.58%, producing a good model for the diagnosis of brain tumour illness. However, these networks also require careful hyperparameter tweaking.

The VGG-16 model, a deep CNN architecture well-known for its efficacy in image recognition tasks, has been applied effectively for brain tumor detection in medical imaging [21]. The VGG-16 model has been modified for the segmentation and classification of brain tumours from MRI data. Its deep architecture and use of small (3×3) convolution filters allow for a robust feature extraction. Using a collection of MRI pictures, the model was trained to accurately recognize and categorize several kinds of brain tumours. By using this method, it is shown that VGG-16 can manage the complexity of medical imaging datasets. The comparison analysis presented in the research demonstrates the VGG-16 model's improved performance over other deep learning techniques as well as standard approaches. The efficacy of VGG-16 was demonstrated in terms of recall, precision, and total accuracy, highlighting its clinical applicability. Although deep learning models such as VGG-16 have many benefits, their application in clinical settings is not without its difficulties. These include managing big datasets and the requirement for substantial processing resources. Subsequent avenues for research could concentrate on refining these models to increase their effectiveness and flexibility in practical medical environments.

AlexNet was introduced to explore large-scale image classification through deep learning. AlexNet has a main advantage in terms of high-performance accuracy due to its deep architecture and uses ReLU as its activation function, which in turn helped to overcome the vanishing gradient problem [12]. However, it comes with high computation costs, and it is more complex hence making not ideal for resource-restricted settings.

VGG16, a simpler yet effective model for feature extraction, utilizes small convolutional filters to enhance performance in image recognition tasks [22]. Its weakness comes from having too many parameters making it prone to overfitting and it requires heavy computation.

For improved brain tumour segmentation from MRI images, Inception modules were incorporated into the U-Net architecture to enhance feature extraction and segmentation accuracy [23]. The study makes use of convolutional neural networks (CNNs) to extract intricate features from big datasets. The U-Net framework is a popular option for medical picture segmentation because of its effective structure for gathering both contextual and localized information. A unique architecture was suggested that incorporates Inception

modules into the U-Net framework. Due to their reputation for handling data at numerous scales at once, Inception modules are especially useful for accurately representing the varied textures and shapes of brain tumours. The Multimodal Brain Tumour Image Segmentation (BRATS) dataset, a well-known industry benchmark, is used in the study. Due to the multi-scale processing capabilities of Inception modules, the integration of these modules into U-Net for brain tumour segmentation provides a substantial leap in medical imaging by providing a more nuanced feature extraction. More accurate and dependable segmentation outputs are essential for later medical operations and treatments, and this progress may result in them. One of the drawbacks would be that adding Inception modules to the U-Net architecture makes the model more complex. Higher computing demands during training and inference may arise from this, and not all clinical settings, especially those with restricted computational resources may be able to support them.

Inception architectures reduce computational complexity by using parallel convolution filters of multiple sizes, enabling efficient multi-scale feature extraction [24]. These architectures have a great capability to extract multi-scale features though it lacks robustness against the hyper-parameter setting and it consumes a lot of computational power.

ResNet addresses the vanishing gradient problem by using residual learning, which enables the construction of deep networks without sacrificing accuracy [25]. ResNet effectively trains very deep networks but still has problems with computational complexities.

EfficientNet utilizes a compound coefficient to effectively scale network dimensions, optimizing both performance and computational efficiency [26]. While being more efficient in terms of complexity, this architecture achieved state-of-the-art accuracy on many benchmarks. The architecture's drawback came with issues of complexity for implementation in medical areas even though its scaling of depth, width, and resolution made it highly suitable.

Various Machine learning algorithms (SVM, K-Nearest Neighbour (KNN), Random Forest) and Deep learning algorithms (Xception, InceptionV3, VGG19, DenseNet201) were trained and tested for classifying gliomas, meningiomas, and pituitary tumours using MRI images from the Figshare dataset [27]. In order to improve the classification accuracy of tumours, a modified Xception model was developed. Categorization performance was measured using precision, recall, F1 score and accuracy. The experimental results confirmed that the deep learning outperformed classical machine learning models and reported results; in particular the modified Xception model provided the highest overall accuracy level exceeding 98%.

The ability of deep learning and traditional machine learning algorithms were evaluated for categorization of brain tumours based on MRI images [28]. SVM, Random Forest, K-Nearest Neighbour (KNN), and six other algorithms were compared with two deep learning models, namely 2D CNN and a convolutional auto-encoder. Both deep learning models achieved high accuracy (2D CNN: 96.47%, auto-encoder: 95.63%) with CNN being slightly better. The findings suggest good applicability of deep learning models in medical imaging tasks, especially because of feature extraction. However, despite the increase in complexity of the auto-encoder no substantial improvements were seen in

the performance of this procedure over the standard CNN model and thus the trade-off between model complexity and actual performance becomes questionable. The traditional machine learning models on the other hand yielded comparatively lower accuracy and that proves that they are not so efficient for such type of work. The strength of the study is the ability to compare and contrast multiple techniques to give insight into which models are more effective. However, the work does not go into much detail on how different hyperparameters affect the model, thus opening the door for future work on the optimization of hyperparameters.

An analysis of nine different pre-trained CNN architectures, including AlexNet, VGG, GoogLeNet, and ResNet, demonstrates their applicability in classifying three types of brain tumors—gliomas, meningiomas, and pituitary tumors—using MRI scans [16]. Good results were reported when using the VGG16 model where the classification accuracy was 98.71%. The paper's strength is in systematically investigating transfer learning, especially for small medical datasets where overfitting and training time can be minimized. It also argued that simplified models like shallower architectures: AlexNet, VGG16 performed significantly better than complex architectures: ResNet- Inception-v2 in terms of accuracy and training time which reveal that high complex models might not always be beneficial for small datasets. It is proved that transfer learning is a feasible solution for the classification of brain tumours, which requires a low amount of computational power while achieving satisfactory performance.

The application of deep learning techniques, particularly convolutional neural networks (CNNs), in various medical image analysis tasks were examined [29]. The corpus is defined for numerous tasks like classification, detection, segmentation, and image synthesis applicable to neuro, retinal, and breast image analysis. It was stated that deep learning approaches are popular in the literature due to their effectiveness in extracting hierarchical features from raw data while generally surpassing traditional machine learning methods commonly used with hand-defined features. The main limitations of deep learning in the context of medical imaging, are the requirement of large, annotated datasets, overfitting in small medical datasets, and the interpretability of the deep learning approach. Overall, a good foundation is formed whereby the study offers guidelines for next steps and points to a number of significant gaps – primarily, data standardization and model interpretability.

A hybrid approach that integrates deep learning algorithms, such as AlexNet and ResNet-18, with a machine learning algorithm (SVM) was developed for classifying brain tumors from MRI images [17]. The highest results were obtained with the AlexNet+SVM combined model with accuracy 95,1%, sensitivity 95,25%, and specificity 98,50%. Hybrid techniques have shown to enhance classification speed while lowering the expense of computation in contrast to deep learning models on their own.

Several pre-trained CNN models, including EfficientNetB0, ResNet50, and InceptionV3, were evaluated for brain tumor classification using three datasets: Kaggle, BraTS 2018, and a real-time dataset from PGIMER [30]. Based on these findings, it is clear that applying hyperparameter tuning to EfficientNetB0 yields an adequate performance, reaching the highest accuracy of 97.32% on real-time data and 96.48% accuracy on BraTS 2018, confirming that transfer learning is beneficial for MRI-based tumour classification. In terms of model comparisons, while the study was extensive, there were many

aspects that were not explored in depth: transfer learning, overfitting, and the differences between various hyperparameters optimization methods across different datasets.

The applicability of deep learning and traditional machine learning in medical imaging has been justified through the literature. An advantage of CNN models as opposed to others, is that they are able to automatically identify hierarchical features of the input data for analysis. The ResNet and EfficientNet architectures are robust models in terms of complexity due to their network depth and efficiency. Unlike CNNs, the traditional machine learning models, including the Support Vector Machine (SVM), Random Forest, and Logistic Regression heavily depend on hand-designed features. However, these models are computationally less demanding and more appropriate for settings where minimal resources are required.

Limitations come in with high computational cost that is inherent in training and applying deep CNNs. The issue of overfitting still persists, despite regularization, which is more significant for limited samples of annotated medical data. The use of architectures like DenseNet and EfficientNet are not quite as suitable for clinical applications due to their sophisticated architectures entailing a great deal of expertise and computing capabilities. CNNs “black box” characteristic, poses a significant concern for medical diagnosis and treatment, where it is important to understand the reasoning and selection processes. The current research intends to go further by determining and accessing the performance of VGG16, Inceptionv3, ResNet50, EfficientNetB0, SVM, Random Forest, and Logistic Regression to detect brain tumours from MRI scans.

3 Methods and Techniques

In this section, the approaches and methodologies applied to build and assess Machine Learning and Deep Learning models to detect brain tumours based on MRI images are described. The methodology includes dataset specification, data preprocessing, model selection, and the design of both traditional and pre-trained models. Each section is crucial for achieving high robustness and reliability in this medical diagnostic task. The flowchart of the detection models is illustrated in Fig. 1.

3.1 Dataset

The dataset, “Br35H: Brain Tumour Detection 2020” by Ahmed Hamada, was obtained from Kaggle. It contains three folders: yes, no, and pred, with 3060 brain MRI images. The yes folder includes 1500 images categorised as tumorous, while the no folder contains 1500 non-tumorous images. The dataset provides labelled images in PNG format, using binary annotations to indicate tumour presence: 0 for the no folder (no tumour) and 1 for the yes folder (tumour). Figures 2 and 3 show the classification of these images. This structure simplifies the generation of labels compatible with binary classifiers and ensures uniformity.

Organising images into class-based folders not only aids in labelling but also facilitates further processes, such as splitting the data into training, validation, and testing sets. By using a publicly available dataset, this research allows for independent verification of evaluation methods, enhancing its credibility.

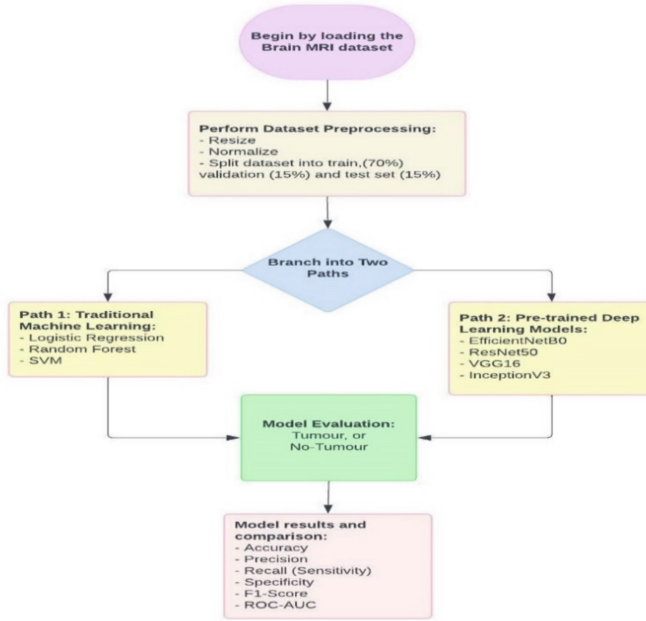


Fig. 1. A flowchart of the detection process

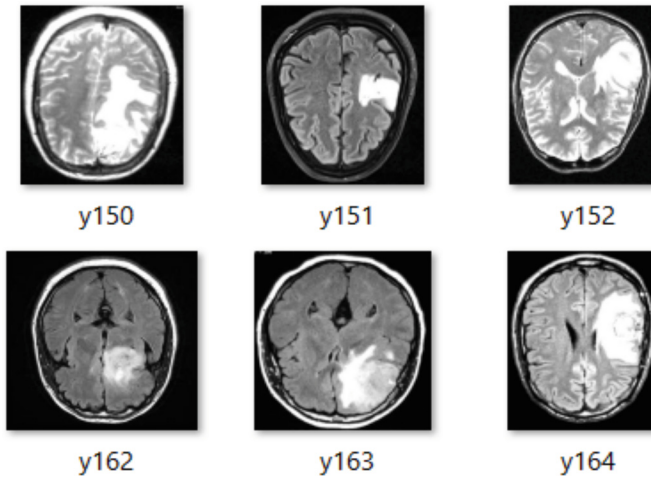


Fig. 2. Sample of dataset images that include tumours

3.2 Preprocessing

Before feeding the images into the models, preprocessing is essential to enhance prediction accuracy. The following operations were performed:

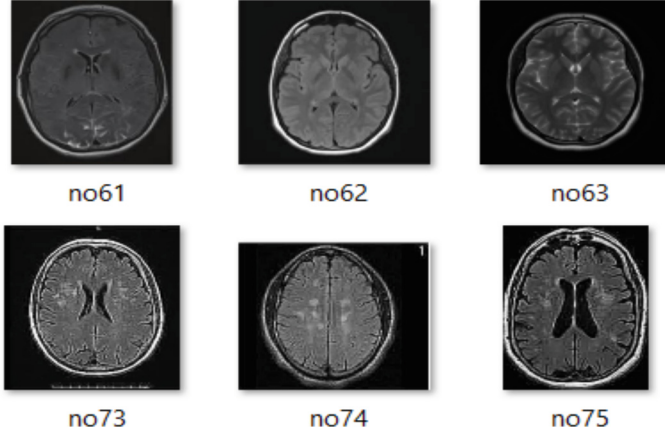


Fig. 3. Sample of dataset images that do not include tumours

- **Image Loading and Resizing:** MRI images were loaded using TensorFlow's `image.load_img()` function, then cropped and resized to 224×224 pixels. This resizing standardises input sizes, which is vital for Deep Learning models like EfficientNet, ResNet, and VGG16.
- **Conversion to NumPy Arrays:** The resized images were converted into NumPy arrays using `image.img_to_array()`, ensuring compatibility with TensorFlow and Keras.
- **Normalisation:** Pixel intensities were normalised by dividing by 255, scaling values to a range of 0 to 1. This step mitigates gradient instability during backpropagation and aligns with the expectations of pre-trained models that typically use normalised pixel values.
- **Data Splitting:** The dataset was divided into three subsets: 70% for training and 15% each for validation and testing. This split avoids order-based bias and ensures a balanced representation of tumour and non-tumour samples, crucial for effective model training.

These preprocessing steps standardise input sizes, stabilise training processes, and enhance generalisation on unseen data—all key objectives in medical image classification.

3.3 Model Implementations

In this section, the various models implemented in this study and their experimental setups are discussed. The structure of the ResNet50 architecture is shown in Fig. 4.

(a) Traditional Machine Learning Models.

Logistic Regression: Logistic Regression was adopted as the baseline model. The pixel values of each MRI scan were flattened and input into the logistic regression classifier. The logistic function is defined by equation (1):

$$h_{\theta}(x) = \frac{1}{1 + e^{-(\theta^T x)}} \quad (1)$$

where $h\theta()$ estimates the probability that an image contains a tumour, θ are the model parameters, and xxx is the input feature vector (pixel values).

Logistic Regression is suitable for binary classification and provides interpretability, making it a valuable tool in medical diagnostics.

Random Forest: This ensemble learning technique builds multiple decision trees and outputs the majority class. The decision function for a tree at a node t is based on Gini Impurity:

$$G(t) = 1 - \sum_{i=1}^C p_i^2 \quad (2)$$

where p_i is the ratio of instances of class i at node t and C is the number of classes (2 in this case).

Random Forests reduce overfitting risks and provide feature importance scores, helping to identify significant areas in the images.

SVM (Support Vector Machine): SVM identifies the optimal hyperplane to separate classes in high-dimensional space. Initially in 2D (224×224), the input data was flattened into one dimensional array. An RBF kernel was employed to classify non-linear relationships effectively. SVMs excel in high-dimensional data scenarios, making them suitable for this task.

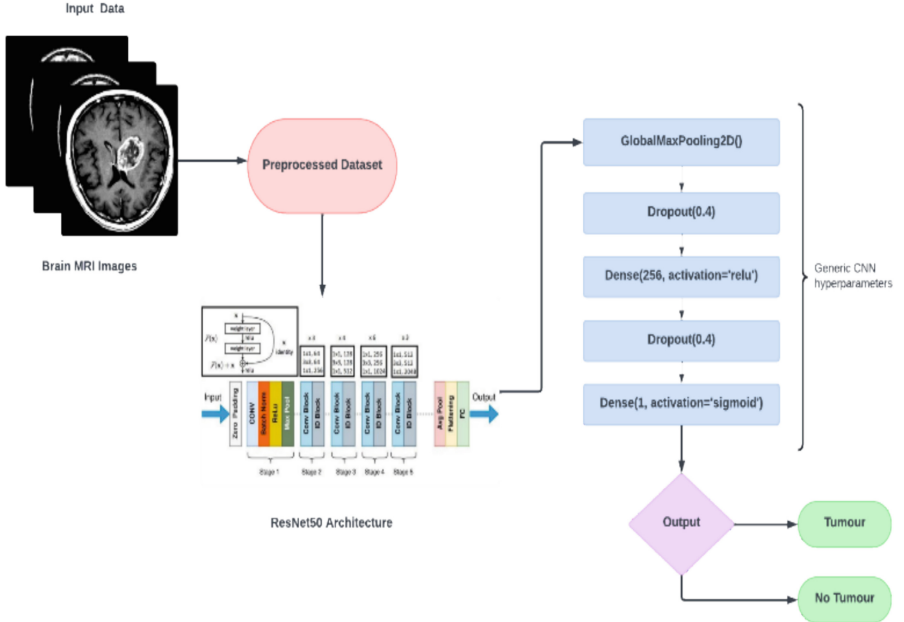


Fig. 4. ResNet50 architecture [31] with added hyperparameters

(b) Pre-trained CNN architecture.

Various CNN architectures were utilised, including EfficientNetB0, ResNet50, VGG16, and InceptionV3. These pre-trained models are well-recognised for their efficiency in image classification tasks and leverage transfer learning using weights from ImageNet.

A transfer learning strategy was applied by freezing the weights of the initial convolutional layers, which capture generalisable features, and training the later layers for the specific classification task. Each model's output layer was modified to include a dense layer with a sigmoid activation function for binary classification.

A generic CNN model was also developed to establish a baseline for comparison. Images were resized to 224×224 pixels. Dropout layers were incorporated to mitigate overfitting, adjusting rates based on the model used. Hyperparameters, including the Adam optimiser with an adjustable learning rate set to 0.0001, were systematically tuned using the grid search process. This made it possible to compare different learning rates (0.0001 and 0.001), dropout rates (0.3, 0.4 and 0.5) and different batch sizes (8, 16 and 32) to establish which set of parameters provided the best convergence rate and model accuracy. The models were trained for 100 epochs, employing early stopping based on validation loss with patience of 10 epochs to prevent overfitting. The binary cross-entropy loss function was used for binary classification.

EfficientNetB0: Known for its scaling depth, channels, and input size efficiency, EfficientNetB0 incorporates dropout rates of 0.5 and 0.4 to control overfitting, with a batch size of 16.

ResNet50: With 50 layers, ResNet50 leverages residual connections to address the vanishing gradient problem. A dropout rate of 0.4 was implemented to enhance generalisation, with a batch size of 32.

VGG16: A simpler architecture with 16 layers, VGG16 employs consecutive convolutional layers. Dropout rates of 0.4 were applied after critical layers to prevent overfitting, and a batch size of 32 was used for effective training.

InceptionV3: This model features Inception modules for simultaneous complex convolution processes, excelling in capturing features at various scales. A dropout rate of 0.5 was applied after the GlobalMaxPooling2D layer and a dropout rate of 0.4 was applied before the last dense layer to mitigate overfitting, with a batch size of 32. Combining pre-trained models with a generic CNN establishes a robust experimental framework for this research. Figure 5 illustrates a summary of the model hyperparameters.

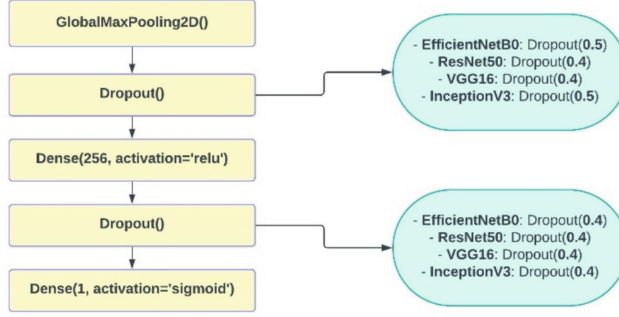


Fig. 5. A flowchart of the Generic hyperparameters used for each deep learning model and their respective dropout rates.

4 Results and Discussion

4.1 Evaluation Criteria

The evaluation metrics used to assess the models are defined by the Eqs. (3)–(7) and Fig. 6 showing a confusion matrix plot.

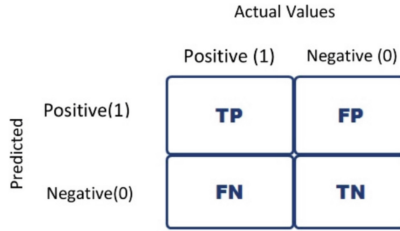


Fig. 6. A Confusion Matrix Plot [35].

- **TP (True Positive):** Correctly predicted as a tumour.
- **FP (False Positive):** Incorrectly predicted as a tumour.
- **FN (False Negative):** Incorrectly predicted as no tumour.
- **TN (True Negative):** Correctly predicted as no tumour.

- 1) **Accuracy:** measures the overall correctness of the model. Higher accuracy indicates better performance [35]. This is given by Eq. (3):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

- 2) **Precision:** indicates how often the model correctly predicts positives, thus reducing false positives [35]. Higher precision means fewer incorrect positive predictions. This is given by Eq. (4):

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4)$$

- 3) **Recall (Sensitivity)**: measures the model's ability to correctly classify true positives [35]. Higher recall signifies fewer false negatives (see Eq. (5)):

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

- 4) **Specificity**: assesses the model's ability to classify true negatives correctly. Similar to recall, it applies to the negative class; higher specificity indicates fewer false positives, signifying accurate identification of negative cases [35]. This is given by Eq. (6):

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (6)$$

- 5) **F1-Score**: balances the precision and recall metrics [35]. A higher F1 score indicates a better balance, especially in cases of uneven class distribution. This is given by Eq. (7):

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

- 6) **ROC-AUC (Area Under the Curve)**: measures the model's ability to distinguish between positive and negative classes. A higher AUC indicates better model performance, with a score of 1 representing excellent classification [35].

4.2 Experimental Results

The experimental results gotten can be summarised in Tables 1 and 2. Table 1 summarises the performance metrics for each Deep Learning and traditional Machine Learning model. Precision values are significant, as misclassifying a patient as having a tumour can lead to severe consequences. Table 2 presents the confusion matrices for each model.

4.3 Discussion

- 1) **Analysis of Results**: Tables 1 and 2 show that Deep Learning models (EfficientNetB0, ResNet50, VGG16, and InceptionV3) consistently outperform traditional Machine Learning models (Logistic Regression, Random Forest, and SVM).

ResNet50 sets the benchmark with an outstanding performance across all evaluation criteria, achieving a precision of 99.11% and recall of 99.55%. The ResNet50 model has more than 23 million parameters (weights and biases within the network's layers), which allows the model to capture finer details of MRI images, which are critical in distinguishing between tumour and non-tumour areas. ResNet50 has performed well due to the presence of residual learning that mitigates the vanishing gradient problem that arises with deeply layered structures in neural networks. These skip connections give the input an opportunity to bypass through some of the layers and be added directly to the output; this preserves the gradient flow and thus make learning with 50 layers possible. It allows the model to learn only the residual, or difference, in each block between the input and output, instead of relearning all the features from scratch. Therefore, ResNet50

effectively understands the patterns in way that it avoids overfitting and hence performs well on unseen data. This is justified through its confusion matrix which reveals only two false positives and one false negative, resulting in a specificity of 99.12% and an impressive ROC-AUC score of 99.98%.

EfficientNetB0 follows closely with precision at 98.67%, recall at 99.11%, and specificity at 98.67%. EfficientNetB0 has over 4 million parameters, specifically its weights and biases, making it more computationally efficient, while still maintaining a good accuracy of 98.89%. This is due to its compound scaling approach, optimally balancing network depth, width, and input image resolution. This efficiency, however, makes it not as good in feature-capturing as deep models, and can potentially miss more complex patterns, as seen by a confusion matrix showing three false positives and two false negatives.

InceptionV3, with over 21 million parameters (comprising its weights and biases), matches ResNet50 with a precision of 99.11% and specificity of 99.12% while maintaining a recall of 99.11%. The Inception modules, which adaptively deal with several filter sizes (1×1 , 3×3 , 5×5), help it embody multi-scale characteristics that are significant for discovering tumours, with different sizes shapes. For this reason, InceptionV3 is highly efficient when it comes to tumour recognition at various scales. Nevertheless, due to its cascaded multi-flow design and a large number of parameters, its performance heavily relies on the optimization of hyperparameters and may overfit slightly on specific patterns of the dataset. This complexity is shown in its confusion matrix indicating two false positives and two false negatives.

VGG16, despite its high recall of 99.55%, has a lower precision of 97.38% and specificity of 97.35%, resulting in six false positives. However, it maintains an overall accuracy of 98.44% and an F1-score of 98.45%. VGG16 has around more than 14 million parameters (primarily weights and biases) and due to sequential layers, which use small (3×3) convolutional filters it is relatively easier to extract features. Due to the conceptual simplicity of the architecture and relatively high number of parameters, the model is sensitive to overfitting. This is probably the cause of the high number of false positives since VGG16 might be highly focused on features that are common with tumours and non-tumours. This in turn can result in redundant feature learning, which can decrease its specificity compared to, for example, ResNet50 due to the lack of residual connections, hence, the large number of parameters.

In contrast, traditional Machine Learning models like Logistic Regression, Random Forest, and SVM exhibit lower performance. Logistic Regression, for instance, achieves a precision of 96.85% and a recall of 95.98%, with nine false negatives. A disadvantage of Logistic Regression is that it defines the decision boundary using a linear equation, and thus cannot identify non-linear characteristics in the MRI data which play a crucial role in discriminating tumour features.

Random Forest's performance is notably poorer, with a precision of 91.88% and a high rate of false positives (19). Even though Random Forest employs ensemble learning, it is confined by the necessity of handcrafted features. The absence of depth and hierarchical learning makes it difficult to deliver the levels of feature detection seen in CNNs.

Table 1. Performance metrics for each model.

Model	Precision	Accuracy	Recall	Specificity	F1-Score	ROC-AUC
EfficientNetB0	98.67	98.89	99.11	98.67	98.89	99.96
ResNet50	99.11	99.33	99.55	99.12	99.33	99.98
VGG16	97.38	98.44	99.55	97.35	98.45	99.88
Inception V3	99.11	99.11	99.11	99.12	99.11	99.99
Logistic Regression	96.85	96.44	95.98	96.90	96.41	99.16
Random Forest	91.88	93.78	95.98	91.59	93.89	97.46
SVM	93.51	94.89	96.43	93.36	94.94	98.35

Table 2. Confusion matrices for each model.

Model	True Negative	False Positive	True Positive	False Negative
EfficientNetB0	223	3	222	2
ResNet50	224	2	223	1
VGG16	220	6	223	1
Inception V3	224	2	222	2
Logistic Regression	219	7	215	9
Random Forest	207	19	215	9
SVM	211	15	216	8

SVM achieves a precision of 93.51%, indicating a lower accuracy than the Deep Learning counterparts. Despite the fact that the RBF kernel in SVM provides capability to capture some degree of non-linearity, it does not have the convolutional layers capacity for hierarchical and spatial information which is critical for MRI image analysis.

2) **Analysis of Accuracy and Loss Curves:** The accuracy and loss curves for the ResNet50 model (see Fig. 7) demonstrate excellent learning and generalisation. Both training and validation accuracies reach high values in early epochs, suggesting effective learning of features from the dataset. Minor oscillations in validation accuracy indicate slight overfitting, but both curves converge towards a similar value, signifying the model’s capacity to learn appropriate patterns. Loss curves drop rapidly, with a brief increase in validation loss likely caused by complex samples- data points from the validation set that display more complicated properties, such as different types of tumours or overlapping features. These samples put pressure on the model, causing misclassifications and a brief increase in the validation loss. However, as the model learns and starts experiencing more complex cases, it corrects its performance, and the validation loss starts to become stable.

The VGG16 model (see Fig. 8) shows a steep increase in training accuracy, nearing 100% within ten epochs, while validation accuracy rises to 80% initially and stabilises

between 97% and 99%. This close relationship indicates good generalisation without significant overfitting. Both training and validation losses decrease sharply, with training loss approaching zero.

InceptionV3 (see Fig. 9) achieves training accuracy close to 100% by epoch 10, with validation accuracy exceeding 95% thereafter. Validation accuracy oscillates, likely due to sample difficulty, while loss curves indicate minimal error. The EfficientNetB0 model (see Fig. 10) shows rapid convergence in both accuracy curves, indicating mastery in predictions with minimal overfitting. Loss curves reflect this, with both training and validation losses nearing zero by epoch 10, suggesting a well-optimised model.

- 3) **Comparison with previous approaches:** The proposed approach demonstrates significant improvement over previous methods, as seen in Table 3. Models like CNN and VGG16 achieved decent accuracy but lacked depth in feature extraction. In contrast, ResNet50 achieves a remarkable accuracy of 99.33%. The comparison highlights the superior generalisation and stability of the proposed model over traditional Machine Learning methods and prior Deep Learning approaches.
- 4) **Challenges with this approach:** Several challenges emerged during the research, particularly overfitting. While models excelled in training data, they struggled with validation. Techniques such as data shuffling, early stopping, and careful hyperparameter tuning were implemented to address this. Restricted GPU access in Google Colab also limited training time and raised concerns about RAM usage, especially for Deep Learning models, which typically require extended training periods.
- 5) **Limitations of this Study:** This study acknowledges several limitations that may impact the results and their generalizability. Although Deep Learning models showed high accuracy, their complexity raises interpretability concerns and the risk of overfitting. Hyperparameter tuning was computationally expensive, and optimal parameters may not transfer easily to other datasets. Limitations in Google Colab, such as GPU hours and RAM constraints, further affected experimentation. Additionally, findings specific to certain tumour types may not generalise to other classifications. Future research should replicate and extend these findings across diverse clinical populations.

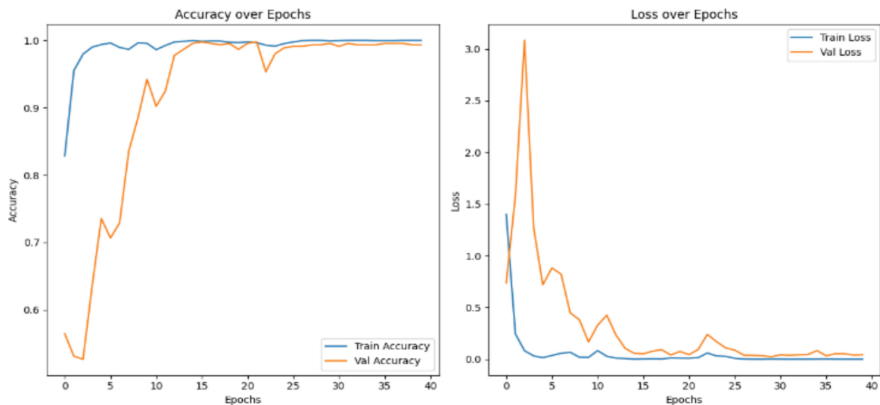


Fig. 7. ResNet50 Accuracy and Loss curves

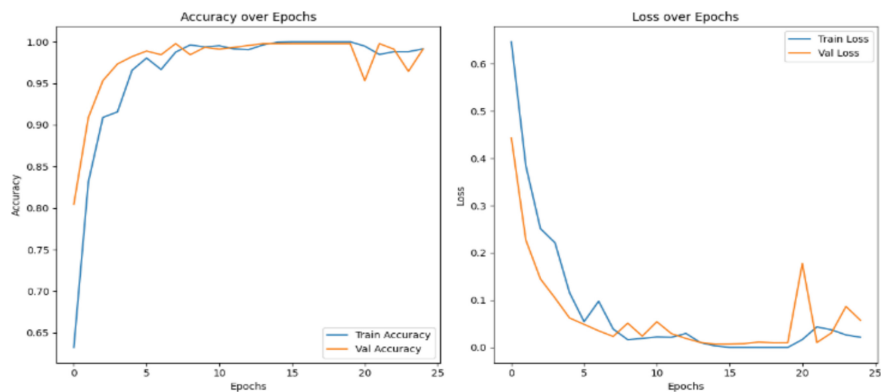


Fig. 8. VGG16 Accuracy and Loss curves

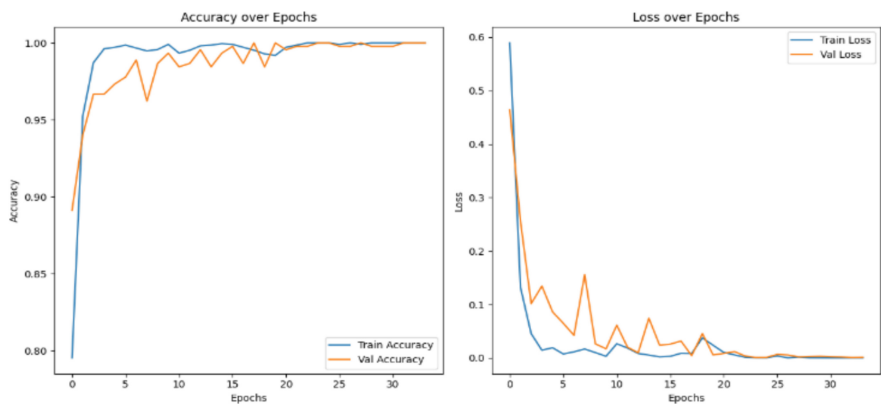


Fig. 9. InceptionV3 Accuracy and Loss curves

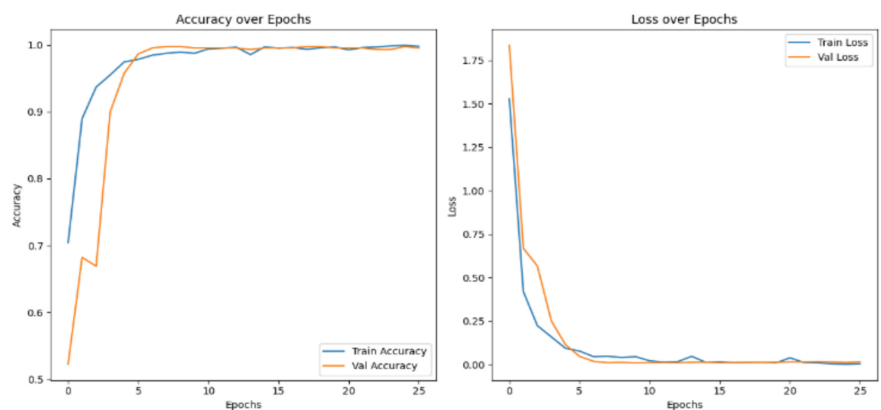


Fig. 10. EfficientNetB0 Accuracy and Loss curves

Table 3. Comparison with previous approaches.

References	Architecture	Precision	Accuracy
Choudhury et al. [32]	CNN	94.87%	96.08%
Togacar et al. [33]	VGG16	89.66%	84.48%
Khan et al. [34]	ResNet-50	87%	89%
Khan et al. [34]	Inception-v3	77%	75%
Gómez-Guzmán et al. [31]	EfficientNetB0	93.12%	90.88%
Gómez-Guzmán et al. [31]	InceptionV3	97.97%	97.12%
Gómez-Guzmán et al. [31]	ResNet50	97.96%	96.97%
Hossain et al. [35]	SVM	93.5%	92.42%
Hossain et al. [35]	Logistic Regression	91.8%	87.88%
Hossain et al. [35]	Random Forest	90.3%	89.39%
Proposed approach	ResNet50	99.11%	99.33%

5 Conclusion and Future Work

This research demonstrates that Deep Learning models outperform traditional Machine Learning methods in predicting brain tumours from MRI scans. Notably, ResNet50 emerged as the most accurate model, achieving a precision of 99.11%, a recall of 99.55%, and an overall accuracy of 99.33%. Other Deep Learning models, such as EfficientNetB0, InceptionV3, and VGG16, also exhibited strong performance, indicating their effectiveness in classifying brain tumours with fewer false positives and false negatives. In contrast, traditional models like Logistic Regression, Random Forest, and Support Vector Machine displayed lower accuracy, precision, and recall and higher rates of false positives. These findings suggest that while classic models can provide reasonable classifications, they are less suited for medical diagnostics, where precision and accuracy are critical. The accuracy and loss curves analysis reinforced the stability of ResNet50 and VGG16, highlighting minimal overfitting and strong generalisation capabilities. Both models quickly learned important features from the dataset, maintaining high accuracy from the outset of training. Therefore, deep Learning models, particularly ResNet50, offer superior accuracy and reliability for detecting brain tumours in medical applications. This advancement can help mitigate misdiagnosis and improve patient outcomes compared to traditional Machine Learning models.

In future, research may focus on integrating the strengths of Deep Learning with traditional Machine Learning techniques to develop a hybrid model, thereby enhancing performance while simultaneously reducing computational costs. Furthermore, there is potential to expand the current model into a multi-class framework, which would facilitate the identification of various complex types of brain tumours. Additionally, the methodology could be adapted for diagnosing other medical conditions, such as lung cancer, thereby broadening the applicability of these models within the medical field.

References

1. Siegel, R.L., Miller, K.D., Jemal, A.: Colorectal cancer statistics. *CA: A Cancer J. Clinic.* **67**(3), 177–193. (2017)
2. Gopal, S.T., Dharuman, G., Nachiappan, S.: A review on a deep learning perspective in brain cancer classification. *Cancers (Basel)* **11**(1) (2019)
3. Gore, D.V., Deshpande, V.: Comparative study of various techniques using deep learning for brain tumor detection. In: 2020 International Conference for Emerging Technology (INCET), pp. 1–4. (2020)
4. Deangelis, L.M.: Brain tumors. *New England J. Med.* **344**(2) (2001)
5. Chatterjee, S., Saini, B.S., Gupta, S.: Classification of brain tumours in MR images using deep spatiotemporal models. *Sci. Rep.* **12**(1), 1505 (2022)
6. Fink, J.R., Muzi, M., Peck, M., Krohn, K.A.: Multimodality brain tumor imaging: MR imaging, PET, and PET/MR imaging. *J. Nucl. Med.* **56**(10), 1554–1561 (2015)
7. Kaur, T., Saini, B.S., Gupta, S.: Quantitative metric for MR brain tumour grade classification using sample space density measure of analytic intrinsic mode function representation. *IET Image Proc.* **11**(8), 620–632 (2017)
8. Usman, K., Rajpoot, K.: Brain tumor classification from multi-modality MRI using wavelets and machine learning. *Pattern Anal. Appl.* **20**(3), 871–881 (2017)
9. Anitha, V., Murugavalli, S.: Brain tumour classification using two-tier classifier with adaptive segmentation technique. *IET Comput. Vision* **10**(1), 9–17 (2016)
10. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 47–52 (1998)
11. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S.: ImageNet large scale visual recognition challenge. *Int. J. Comput. Vision* **115**(3), 211–252 (2015)
12. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: *Proceedings of the Advances in Neural Information Processing Systems*, pp. 1097–1105 (2012)
13. NHS. Brain tumours. National Health Service. <https://www.nhs.uk/conditions/brain-tumours/>. Accessed 20 Nov 2024
14. Tiwari, P., Srivastava, S., Pant, M., Virmani, J.: CNN-based multiclass brain tumor detection using medical imaging. *Comput. Intell. Neurosci.* **2022**, 1830010 (2022)
15. Kouli, O., Hassane, A., Barakat, D., Kouli, T., Hossain-Ibrahim, K., Steele, J.D.: Automated brain tumor identification using magnetic resonance imaging: a systematic review and meta-analysis. *Neuro-Oncol. Adv.* **4**(1), vdac081 (2022)
16. Chelghoum, R., Chaker, A., Rebbah, L., Grati, A.: Transfer learning using convolutional neural network architectures for brain tumor classification from MRI images. In: *Artificial Intelligence Applications and Innovations*, pp. 189–200. (2020)
17. Senan, E.M., Almazroi, A.A., Alsharif, W.R., Salem, N.: Early diagnosis of brain tumour MRI images using hybrid techniques between deep and machine learning. *Comput. Math. Methods Med.* **2022**, 8330833 (2022)
18. Zhao, X., Wu, Y., Song, G., Zhang, J.: A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *Med. Image Anal.* **43**, 98–111 (2018)
19. Sajjad, M., Khan, S., Muhammad, K., Wu, W.: Multi-grade brain tumor classification using deep CNN with extensive data augmentation. *J. Comput. Sci.* **30**, 174–182 (2019)
20. Younis, A., Qadri, S.F., Ghafoor, A.: Brain tumor analysis using deep learning and VGG-16 ensembling learning approaches. *Appl. Sci.* **12**(14) (2022)
21. Simonyan, K., Zisserman, A.: Very deep convolution networks for large-scale image recognition. In: *Proceedings of the International Conference on Learning Representations (ICLR)* (2015)

22. Cahall, D.E., Patel, V.M., Chandra, S.: Inception modules enhance brain tumor segmentation. *Front. Comput. Neurosci.* **13**, 44 (2019)
23. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S.: Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9 (2015)
24. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778. (2016)
25. Tan, M., Le, Q.V.: EfficientNet: rethinking model scaling for convolutional neural networks. In: *Proceedings of the International Conference on Machine Learning*, pp. 6105–6114 (2019)
26. Singh, Y.P., Lobiyal, D.K.: A comparative analysis and classification of cancerous brain tumors detection based on classical machine learning and deep transfer learning models. *Multimedia Tools Appl.* **83**(13), 39537–39562 (2023)
27. Saeedi, S., Safaei, A., Shokri, E.: MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. *BMC Med. Inform. Dec. Mak.* **23**(1) (2023)
28. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F.: A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017)
29. Sachdeva, J., Sharma, D., Ahuja, C.K.: Comparative analysis of different deep convolutional neural network architectures for classification of brain tumor on magnetic resonance images. *Arch. Comput. Meth. Eng.* **31**(4), 1959–1978 (2024)
30. Kiyani, I., Javed, R., Ullah, I., Ali, H.: Deep learning-based Glaucoma Network Classification (GNC) using retinal images. *Int. J. Imag. Syst. Technol.* **34** (2023)
31. Gómez-Guzmán, M.A., García-González, A., González-Navarro, F.F., Martínez-Rojas, J.A.: Classifying brain tumors on magnetic resonance imaging by using convolutional neural networks. *Electronics* **12**(4) (2023)
32. Choudhury, C.L., Mahanty, C., Kumar, R., Mishra, B.K.: Brain tumor detection and classification using convolutional neural network and deep neural network. In: *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, pp. 1–4 (2020)
33. Togacar, M., Ergen, B., Comert, Z.: BrainMRNet: brain tumor detection using magnetic resonance images with a novel convolutional neural network model. *Med. Hypoth.* **134**, 109531 (2020)
34. Khan, H.A., Khan, A.A., Javed, S.: Brain tumor classification in MRI image using convolutional neural network. *Math. Biosci. Eng.* **17**(5), 6203–6216 (2020)
35. Hossain, T., Shishir, F.S., Ashraf, M., Nasim, M.A.A., Shah, F.M.: Brain tumor detection using convolutional neural network. In: *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, pp. 1–6 (2029)



Prostate Cancer Detection and Classification Using Deep Learning Pretrained Model

Olushola Olawuyi^(✉) 

School of Mathematics, Statistics and Computer Science University of
KwaZulu-Natal, Durban, South Africa
221119659@stu.ukzn.ac.za

Abstract. Medical research indicates that prostate cancer is among the deadliest types of cancer. Detecting prostate cancer (PCa) early greatly impacts how effectively the tumor can be treated. Artificial intelligence (AI) is advancing the field of digital pathology, enabling researchers to analyze large datasets and diagnose prostate cancer tumors more quickly and accurately. One method for early and rapid detection of prostate cancer is through image analysis software that uses machine learning or deep learning algorithms. This study evaluates the effectiveness of two convolutional neural network (CNN) algorithms for detecting prostate cancer. Python programming, along with Keras and TensorFlow frameworks, are utilized to implement two deep learning models: InceptionV3 and ResNet50. These models were evaluated using the MyDataset4 prostate magnetic resonance images (MRI) dataset from the Kaggle platform, which contains both cancerous and non-cancerous images. Both models have been applied to various medical imaging diagnostic tasks and have shown outstanding performance. The models are assessed based on accuracy, F1 score, recall, and precision to gauge their performance. Test results indicate that ResNet50 achieved the highest accuracy at 93%, while InceptionV3 reached an accuracy of 91%.

Keywords: Deep learning · Prostate cancer · InceptionV3 model · CNN model · Resnet50 model and Magnetic Resonance Imaging (MRI)

1 Introduction

Cancer occurs when a body cell develops abnormally and grows uncontrollably. In a healthy body, cells typically grow and multiply as needed to replace old or damaged cells; however, this controlled process goes awry in cancer [31]. According to statistics from the American Cancer Society on prostate cancer risk, approximately 1 in every 8 men may undergo prostate cancer treatment at least once in their lifetime. The American Cancer Society also states that prostate cancer is the second leading cause of cancer death among American men, following lung cancer, which is another common cause of death in men. Their projections for 2024 indicate that 1 in every 44 men will succumb to

prostate cancer. Nonetheless, a man's risk of developing prostate cancer can be influenced by factors such as age, ethnicity, and other individual characteristics [32].

In treating this health challenge, prostate biopsies are used to diagnose cancer; however, a problem is that results are not obtained quickly, as the test takes one to three days to complete, or even longer in some cases [10, 20]. Additionally, measuring prostate-specific antigen (PSA) levels alone is insufficient to identify prostate cancer accurately. Therefore, to reduce false positives and frequent errors made by doctors, reliable methods like deep learning algorithms can be employed to detect prostate cancer [10].

Current research indicates that artificial intelligence systems can assist with pathological evaluation for PCa detection and categorization using pathology imaging. Additionally, recent developments in deep learning applications suggest that the performance of PCa grading algorithms may still need improvement [17, 22]. Deep learning methods and computational pathology have demonstrated the capability to achieve expert-level pathology diagnosis [7, 18]. To increase accuracy and reduce costs, artificial intelligence (AI) techniques like deep learning (DL) are being applied to detect PCa. Medical imaging, such as MRI and CT scan data, is analyzed to determine patient characteristics, including race, age, family history, and lifestyle habits. Using a deep learning model for PCa identification can help reduce the costs associated with biopsies and other tests by providing more accurate outcomes compared to conventional methods [35]. Furthermore, deep learning algorithms can handle Gleason scoring with performance comparable to that of medical experts, according to several studies [6]. Deep learning algorithms can handle Gleason scoring with better performance like a medical expert treatment, according to several studies [3, 5, 21, 22, 29, 33].

Deep learning (DL) has demonstrated a sensitivity of up to 95%, making it more accurate than conventional approaches for detecting prostate cancer [35]. It has also proven to be faster than traditional techniques, which require several minutes to identify prostate cancer in each image, providing both speed and accuracy [35]. The advent of artificial intelligence methods has made it possible to identify diseases from medical images, which has significantly promoted its use in medicine today [12]. Finally, the introduction of convolutional neural networks (CNN) and deep learning models has simplified information extraction from images, contributing to high accuracy rates.

1.1 The Following Summarises the Study Contributions:

The primary goal of this study is to develop a solution that will aid in addressing prostate cancer challenges. Therefore:

- We implement two deep-learning pre-trained models for the detection and classification of benign and malignant prostate cancer tumors using an MRI image dataset.
- We evaluate the performance of both models

2 Literature Review

Approximately one million new cases of prostate cancer are estimated to be diagnosed annually worldwide, posing a significant health concern for men [13]. Although deep learning algorithms have been used for prostate cancer diagnosis over the past few years [28,37,39]. A study conducted by [30] elucidates the significance of timely identification and precise treatment of PCa, along with the constraints of existing diagnostic techniques. When their automated solution was tested, it demonstrated high accuracy in identifying cancer spots on prostate biopsy images. The researchers concluded that the accuracy and efficiency of prostate cancer diagnosis could be improved based on the performance of their system [24].

Azizi et al. [4] used a Long Short-Term Memory (LSTM) network to achieve the highest accuracy in distinguishing cancerous tissue from benign tissue in the prostate. The algorithms were presented for an in-depth analysis of LSTM networks. The dataset consisted of 255 prostate biopsy cores from 157 patients, achieving an accuracy of 0.92. Additionally, research conducted by [27], used single and multi-institutional image datasets to assess the precision of a machine learning method in prostate MRI lesion classification.

Furthermore, research conducted by [11] applied deep learning techniques for prostate cancer detection and grading using MRI imaging. A CNN model was used by the authors to analyze the dataset (MRI images) and predict the presence of cancer and its severity. The authors also compared their model to conventional machine learning approaches, and it was proven that convolutional neural networks outperformed the others. The authors concluded from their findings that deep learning architecture is capable of improving prostate cancer diagnosis efficiency and providing better treatment outcomes for patients.

Prostate cancer detection using fully automated computer-aided detection (CAD) was proposed in the study conducted by [31]. The model was used to identify cancerous areas using the CAD technique, which involved a series of continuous processes. Firstly, a voxel classification was performed at several levels using an apparent diffusion coefficient map, with a Hessian-based blob identification approach.

A survey was conducted by [23], providing a systematic overview of deep learning models and methods employed for the management of prostate cancer. A thorough review was carried out on English-language studies spanning two decades, from 2000 to 2021, which were presented in MEDLINE, ClinicalTrials.gov, Scopus, Science Direct, Google Scholar, and Web of Science. A total of 224 studies were discovered in the initial search. After screening, 64 papers related to urology applications were found, of which 24 were exclusively focused on the identification and management of prostate cancer. Research on deep learning applications should continue to grow as deep learning (DL) models evolve. The goal should be to advance these models to the point where they are ready for clinical use. Developing models that can train on encoded images should be a top priority for future research, as this would enable greater accessibility and facilitate data exchange. In a similar vein, the authors [2] proposed a deep-learning

model that attained an AUC of 0.955. The model improved the detection of prostate cancer lesions using magnetic resonance imaging (MRI).

An artificial neural network-based approach called deep learning (DL), a subfield of machine learning, uses data feature extraction [14]. To reduce some inadequacies of artificial neural techniques and enable the construction of end-to-end model predictions, deep learning offers capabilities that now allow computers to automatically identify patterns and incorporate features learned during model or architecture-building procedures [15].

A deep neural network model, XmasNet was used for lesion classification, while 3D U-Net and 3D ResNet-18 models were also employed for image detection and classification [25]. A thorough examination of each issue was conducted using a total of two networks, two cropping strategies, two resampling strategies, two crop sizes, five input sizes, and data augmentations for lesion categorization. The authors indicated that the classification of lesions and prostate detection with segmentation achieved an AUC of 0.870 and a dice score of 0.915. The authors concluded that there was still room for improvement in the segmentation of lesions and the categorization of Prostate Imaging Reporting and Data System (PIRADS) scores (with a dice score of 0.690 and accuracy of 0.664, respectively). Finally, the authors' investigation showed how certain dataset issues could be addressed to enhance performance outcomes.

A deep learning algorithm was employed in a study by [19] to detect prostate cancer in men. 3D MRI images were used for the research, and the model achieved a prediction accuracy of 89% for prostate cancer detection. Prostate cancer was also diagnosed using the CNN algorithm in a study conducted by [36]. The authors achieved approximately 86% accuracy in prostate cancer diagnosis after replicating the suggested procedure.

In a research study by [1], a robust CNN was used with a transfer learning technique. The outcome of their study was compared with several machine learning techniques, such as Decision Tree, Naive Bayes, and SVM. An MRI database of cancer images was used for training the model (GoogleNet) with ML classifiers. Features like SIFT (Scale Invariant Feature Transform), Morphological Texture, Elliptic Fourier Descriptors, and entropy were extracted. During their investigation, various performance evaluation metrics, such as sensitivity, positive predictive value, specificity, negative predictive value, receiver operating characteristic (ROC) curve, and false positive rate, were calculated. In the end, better results were observed using a neural network model (GoogleNet) with the transfer learning method. In conclusion, the authors confirmed that, although good results were obtained from all the machine learning classifiers, such as Decision Tree, Naive Bayes, and SVM, the different kernels showed outstanding performance when using the deep learning (DL) method.

In a study by [38], the authors employed semi-supervised learning alongside a deep learning algorithm to diagnose men with prostate cancer. The model achieved an accuracy rate of 84%. Prostate cancer was diagnosed in research by [8] using an enhanced CNN algorithm. Their accuracy rate for prostate cancer diagnosis after replicating the suggested procedure was 75.1%.

According to [26], in their research on prostate cancer detection, when magnetic resonance imaging is used for prostate cancer biopsy, it consistently yields better results. However, the findings will depend on the capabilities of two pre-trained models, InceptionV3 and ResNet50, to identify and categorize prostate cancer using magnetic resonance imaging (MRI).

Table 1 shows a summary comparison of related studies against the results of the proposed models.

Table 1. The comparison and summary of the related studies and proposed models

Ref	Model	Performance (Accuracy%)
[4]	Long Short-Term Memory (LSTM) Network	92
[25]	XmasNet model	87.0
[19]	Deep learning	89
[38]	Deep learning	84
[36]	CNN	86
[8]	CNN	75.1
Proposed model	InceptionV3	91
Proposed model	ResNet50	93

3 Methods and Techniques

This study used two deep neural network architectures to detect prostate cancer using MRI images: InceptionV3 and ResNet50. The two deep learning models utilize artificial neural networks to predict images. The main aim is to identify the best model of the two that is capable of detecting prostate cancer in magnetic resonance images. In this section, the Inception model and ResNet50 are explained in detail:

3.1 Inception Model

In 2014, Google researchers integrated the Inception architecture to help address the issue of deep neural network processing and computational power constraints. The original Inception architecture, popularly referred to as GoogLeNet, was created to increase deep neural network efficiency while maintaining accuracy and lowering the number of parameters. To extract features from the input image at various scales, the Inception architecture incorporates parallel 1×1 , 3×3 , and 5×5 convolutional filters. The next step in the process is to concatenate the output of each filter before moving on to the next layer. This method reduces the number of parameters while enabling the network to extract both local and global information from the input image. The Inception architecture has been

enhanced and revised multiple times since its introduction. To create a more efficient design with fewer parameters, the researchers released the Inception-v2 architecture in 2015. This architecture used two 3×3 filters in succession instead of five 5×5 convolutional filters [9]. Medical imaging applications of the Inception model, including identifying lung nodules in CT scans and skin cancer in dermatological images, have also been handled with great accuracy; diabetic retinopathy has been identified in retinal fundus images using InceptionV3 [35]. The InceptionV3 architecture was released in 2016 and came with many enhancements, such as batch normalization and a reduction in the number of filters used in each layer. The performance of this model on the ImageNet dataset, which is a benchmark for image classification tasks, demonstrated state-of-the-art results. The introduction of the InceptionV4 architecture incorporated residual connections similar to those in the ResNet design. It was more efficient than earlier iterations of the Inception architecture, and this design produced even better results on the ImageNet dataset. The Inception architecture has greatly enhanced the effectiveness and efficiency of computer vision and deep learning for complex image categorization tasks. The design has evolved to produce progressively more effective and efficient iterations, the most recent of which includes some of the most cutting-edge methods found in deep learning research. Medical imaging applications of the Inception model, including identifying lung nodules in CT scans and skin cancer in dermatological images, have also been handled with great accuracy; diabetic retinopathy has been identified in retinal fundus images using InceptionV3 [9].

3.2 InceptionV3 Model Architecture

InceptionV3 was introduced in 2015; it features 48 layers overall and a lower error rate compared to previous versions. The InceptionV3 model has 48 layers, which is fewer than the Inception V1 and V2 models. However, its performance is quite remarkable. The components that make up the model, compared to the previous Inception models, are shown in Fig. 1. The model is more accurate in terms of performance and requires less computing power.

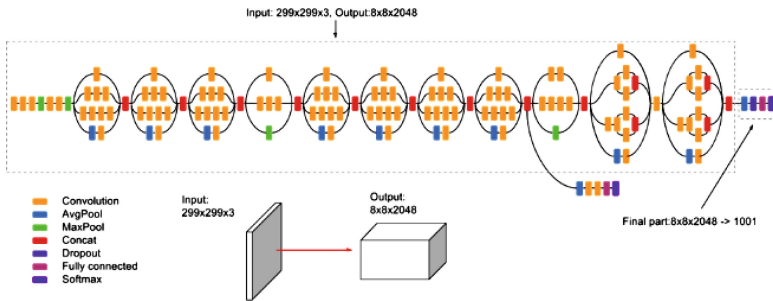


Fig. 1. InceptionV3 Model [34]

3.3 Residual Network (Resnet50)

ResNet50 was developed by Microsoft researchers and introduced in 2015 as a deep neural network architecture. The model is one of the popular versions of the Residual Network (ResNet) architecture, and its 50 layers indicate the depth of the network. This strong image classification model has produced cutting-edge results when trained on large datasets. One of the major strengths of this architecture is its residual connections, which allow the model to learn residual functions that connect the input to the intended output. These residual connections enable the architecture to learn deeper layers without facing the vanishing gradient issue, which was a major challenge in early networks. The four primary components of this network are the convolutional block, convolutional layers, identity block, and fully connected layers. Beginning with the input image, the convolutional layers extract features, which are then processed and transformed by identity and convolutional blocks. The final classification is determined by fully connected layers. After batch normalization and ReLU activation, the convolutional layers remove elements such as edges, shapes, and textures from the input image. Max pooling layers reduce the spatial dimensions of feature maps while retaining important features, following the convolutional layers. The two main building blocks of ResNet50 are the identity and convolutional blocks. The identity block is a straightforward block that adds the input back to the output after passing it through several convolutional layers. This enables the network to learn residual functions, converting the input into the desired output. To reduce the number of filters before the 3×3 convolutional layer, a 1×1 convolutional layer is added to the convolutional block, similar to the identity block. Fully connected layers are the final part of ResNet50, as depicted in Fig. 2, which is responsible for completing the classification task. The final class probabilities are generated by feeding the output of the last fully connected layer into a SoftMax activation function [16].

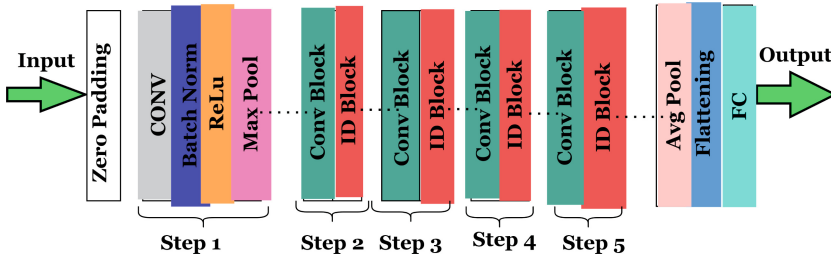


Fig. 2. ResNet50 Architecture [34]

4 Materials and Methods

4.1 The Proposed Model

The proposed architectures for this study are shown in Fig. 3. The model takes MRI images of prostate cancer as input. Afterward, the input image is sent to the preprocessing phase to prepare it for training by the two models, InceptionV3 and ResNet50. The two models were trained and tested with the preprocessed data and then evaluated for performance.

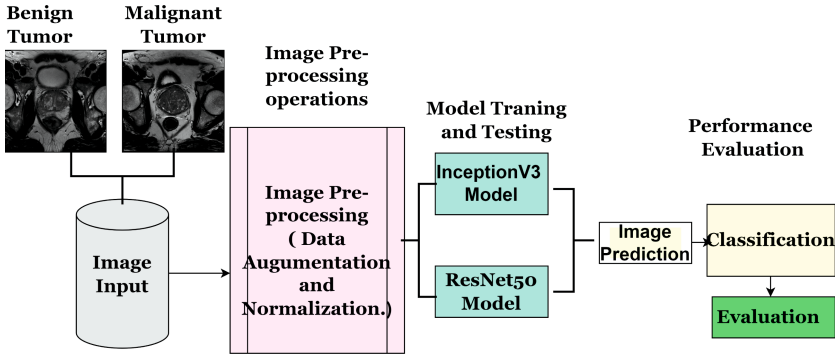


Fig. 3. Proposed architecture

4.2 Dataset

The dataset used for this study consists of a total of 620 magnetic resonance imaging (MRI) medical images. The images have dimensions of 256×256 pixels. When classifying MRIs, it is essential to consider both the intensity of the gray color and the shape or outline of the gray area in the image. The images display cancerous prostate tumors and non-cancerous samples of prostate tissue. The dataset was divided into training, testing, and validation sets as follows: 410 images were used for training, 180 for testing, and 30 for validation. In the training set, there are 244 cancerous MRIs and 166 non-cancerous MRIs; in the testing set, there are 120 cancerous and 60 non-cancerous MRIs; and in the validation set, there are 16 cancerous and 14 non-cancerous MRIs. Dataset visualization can be seen in Fig. 4, and dataset splitting information is provided in Table 2.

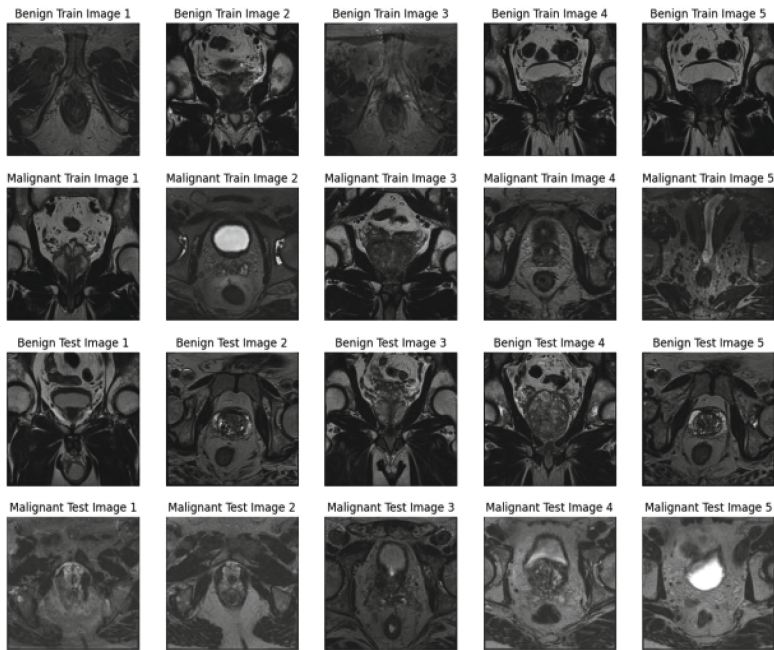


Fig. 4. Dataset Visualization

Table 2. Dataset Splitting into Training, Testing and Validation

Splitting	Number
Number of benign images in the training set	166
Number of malignant images in the training set	244
Number of benign images in the Validation set	14
Number of malignant images in validation	16
Number of benign images in the test set	60
Number of malignant images in the test set	120

4.3 Data Pre-processing

Image augmentation and Normalization were performed as data pre-processing steps in this research. The pixel values in the images were divided by 255 to adjust them to a range of 0 to 1. Normalization helps speed up the training process and results in faster convergence of the models. The augmentation technique ensures that transformations are applied correctly and that the augmented images still represent realistic variations of the original images.

Table 3. Algorithm 1

Pseudocode Based on InceptionV3 and ResNet50 Models
1. Input the Cancer MRI image
2. Initialize parameters
3. Image preprocessing operations
4. Constructing inception V3 model
5. Constructing ResnetV50 model
6. Training both models (inceptionV3 and Resnet50) with labels training data
7. Trained inception V3 performed deeper feature extraction on the input images
8. Trained Resnet50 performed feature extraction on the input images Image prediction by the two models
9. Evaluation of both models for performance result

4.4 Model Training

The two pre-trained models were trained, tested, and validated using an MRI image dataset from Kaggle to determine the presence of benign and malignant tumors, as shown in Table 3. The dataset contains two classes, and a random selection process was used to divide it as follows: 410 images for training, 180 for testing, and 30 for model validation. After each epoch, a callback function was used to plot performance charts for the two models. Both models were set to 100 epochs for training. InceptionV3 converged at the 31st epoch, after which the model showed no further improvement in performance. Additionally, ResNet50 training converged at the 15th epoch, with no further increase in training performance. Validation data was used to test both models at the end of each epoch. Performance metrics for the two models were plotted independently. Both charts displayed validation and training scores for the models. Figures 5 and 6 illustrate the accuracy and loss values for InceptionV3, while Figs. 7 and 8 show the model accuracy and loss values for ResNet50. Validation loss represents the cost function value for cross-validation, while training loss reflects the cost function value for the training data.

4.5 InceptionV3 and ResNet50 Prediction

After the two pre-trained models were sufficiently trained, both were tested to confirm their performance. The results showed that both models could differentiate between benign and malignant prostate tumors. The predictions of the two models are shown in Figs. 9 and 10.

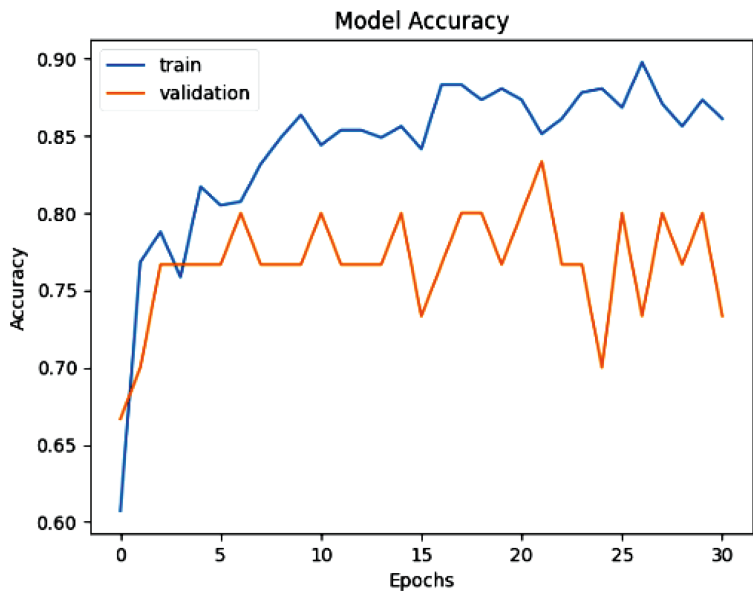


Fig. 5. InceptionV3 Model Accuracy

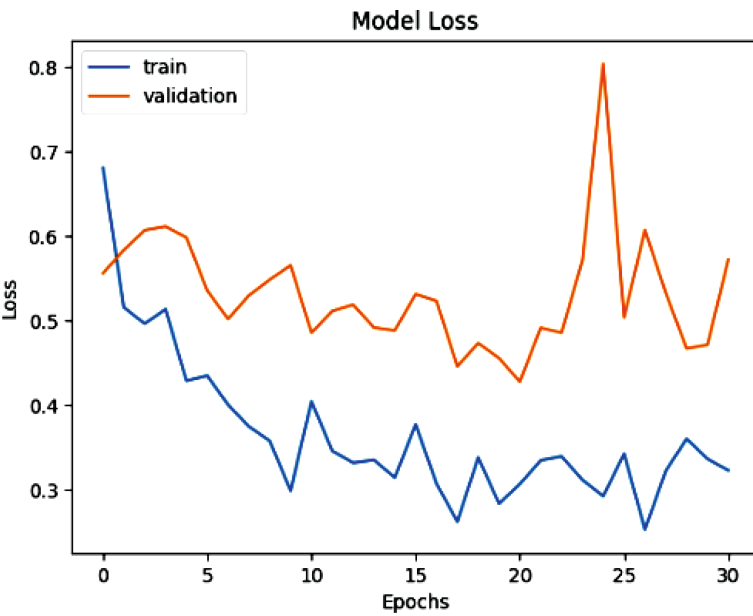


Fig. 6. InceptionV3 Model Loss

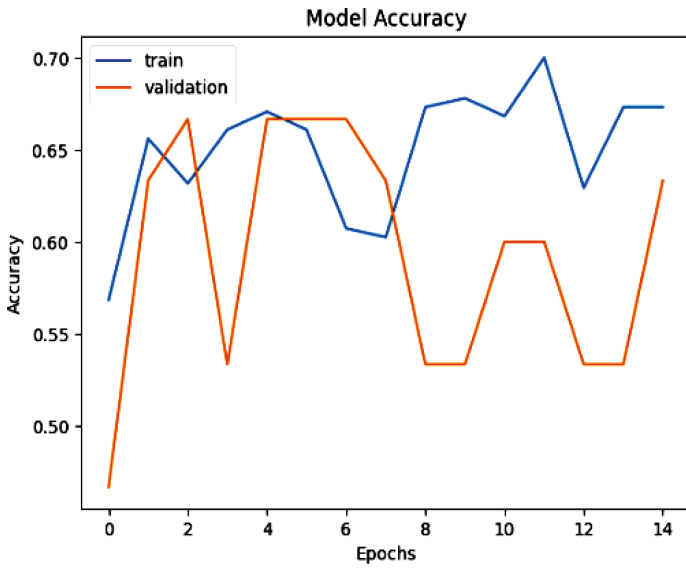


Fig. 7. Resnet50 Model Accuracy

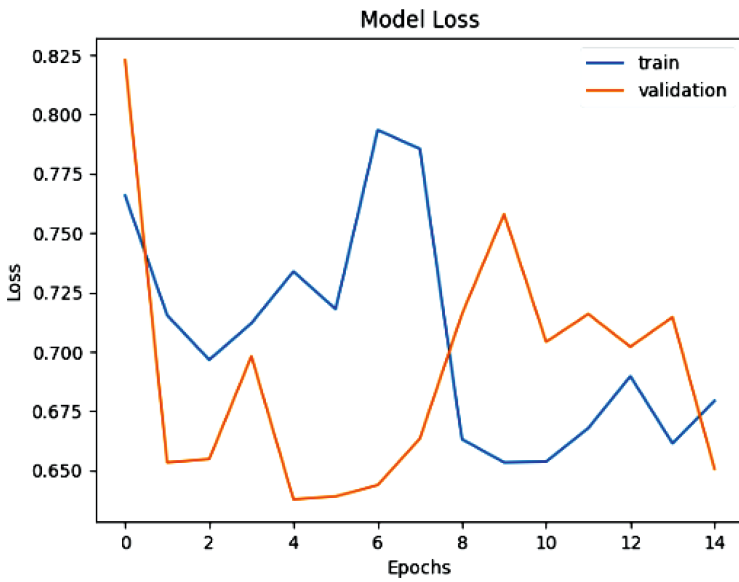


Fig. 8. Resnet50 Model Loss

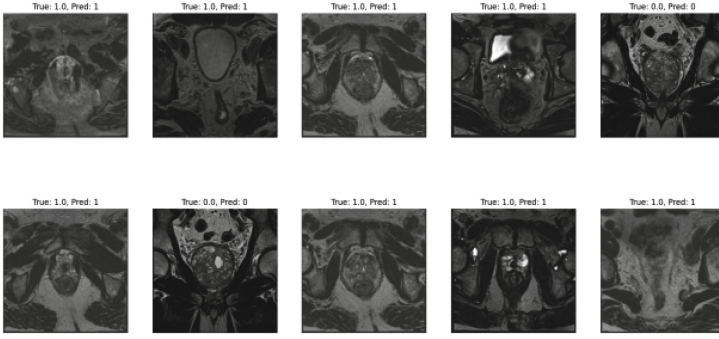


Fig. 9. Predicted images for InceptionV3 model

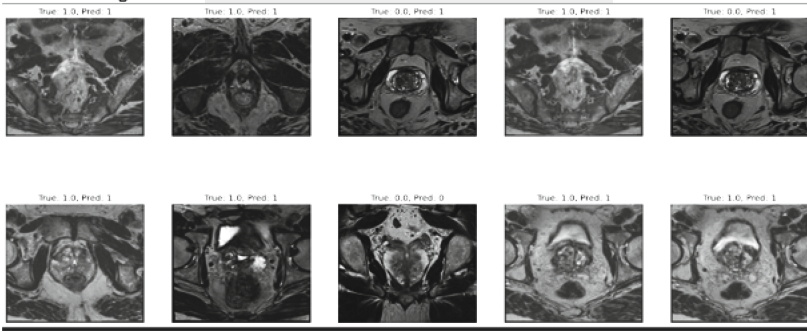


Fig. 10. Predicted images for Resnet50 model

5 Model Evaluation

The performance of the two models was evaluated based on metrics such as recall, accuracy, precision, and F1 score. The formulas used are shown below:

$$ACCURACY = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$PRECISION = \frac{TP}{TP + FP} \quad (2)$$

$$RECALL = \frac{TP}{TP + FN} \quad (3)$$

$$F1SCORE = \frac{2 * Precision * recall}{Precision + Recall} \quad (4)$$

6 Results and Discussion

6.1 Performance of InceptionV3 and ResNet50 for Cancer Detection and Classification

The two deep-learning models were evaluated using four performance metrics. Performance results and parameter settings are presented in Tables 4 and 5. Both InceptionV3 and ResNet50 accurately predicted benign and malignant tumors. InceptionV3 achieved 91% accuracy, 92% precision, 94% recall, and an 93% F1-score, while ResNet50 achieved 93% accuracy, 91% precision, 100% recall, and a 95% F1-score. Binary cross-entropy was used as the loss function for model evaluation during training. The loss value was also compared between the two models during training. When the loss value was higher, the error rate was also higher. Initially, the loss value for InceptionV3 was 0.8, which decreased gradually until it got to 0.28 as training progressed; similarly, ResNet50 initial loss value was 0.72 and also gradually reduced to 0.46.

The results are shown in Table 4, parameter settings in Fig. 5, and metric performance representation in Figs. 11 and 12.

Table 4. Outcome result of the two models.

Model	Accuracy	Precision	Recall	F1-score	Loss
InceptionV3	0.91	0.92	0.94	0.93	0.25
ResNet50	0.93	0.91	1.00	0.95	0.46

Table 5. Parameters Settings

Parameter	Value
Activation function for Input layer	Relu
Activation function for Output layer	Sigmoid
Learning Rate	0.001
Batch size	3
Loss function	binary cross-entropy
Optimizer	Adam
Number of classes	2

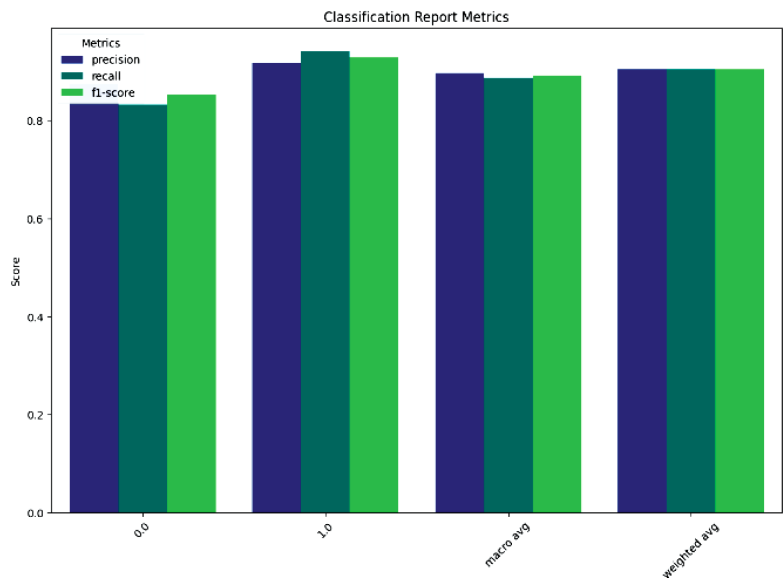


Fig. 11. Performance Metrics for InceptionV3 Model

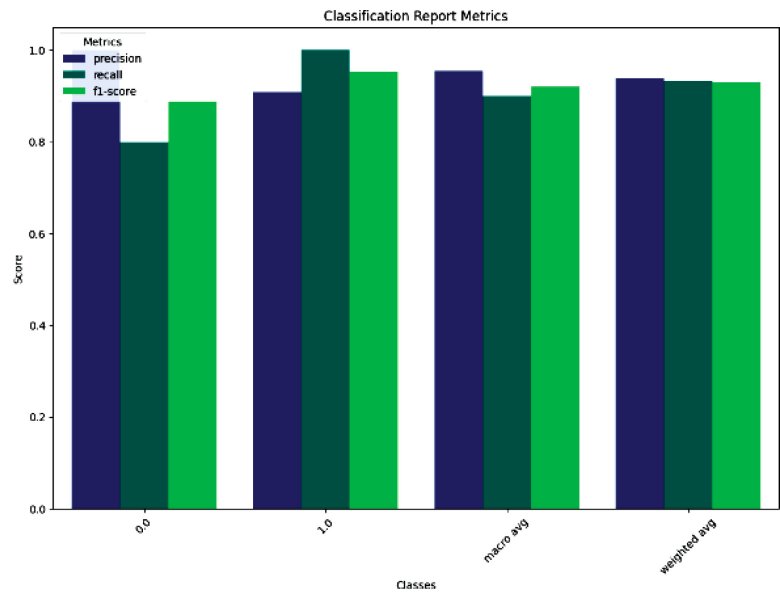


Fig. 12. Performance Metrics for ResNet50 Model

7 Conclusion

This study has demonstrated the ability of the deep learning models InceptionV3 and ResNet50 to detect and classify prostate cancer. The models were evaluated using 620 MRI cancer images, divided into 410 for training, 180 for testing, and 30 for validation. Both models performed well, exceeding a 50% success rate. The ResNet50 model demonstrated superior performance, achieving 93% accuracy, while InceptionV3 reached 91% accuracy. Increasing the number of epochs and dataset size during training can improve outcomes in deep-learning models. Exploring new algorithms could further enhance model performance. The findings suggest that incorporating deep learning models into medical diagnostics can provide a reliable and effective approach for identifying and categorizing prostate cancer, ultimately improving diagnostic accuracy and patient outcomes.

References

1. Abbasi, A.A., et al.: Detecting prostate cancer using deep learning convolution neural network with transfer learning approach. *Cogn. Neurodyn.* **14**(4), 523–533 (2020)
2. Alkadi, R., Taher, F., El-baz, A., Werghi, N.: A deep learning-based approach for the detection and localization of prostate cancer in T2 magnetic resonance images. *J. Digit. Imag.* **32**(5), 793–807 (2019)
3. Arvaniti, E., et al.: Automated Gleason grading of prostate cancer tissue microarrays via deep learning. *Sci. Rep.* **8**(1), 12054 (2018). <https://www.nature.com/articles/s41598-018-30535-1>
4. Azizi, S., et al.: Deep recurrent neural networks for prostate cancer detection: analysis of temporal enhanced ultrasound. *IEEE Trans. Med. Imag.* **37**(12), 2695–2703 (2018)
5. Bulten, W., et al.: Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol.* **21**(2), 233–241 (2020). [https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045\(19\)30739-9/abstract](https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(19)30739-9/abstract)
6. Bygari, R., Rithesh, K., Ambesange, S., Koolagudi, S.G.: Prostate cancer grading using multistage deep neural networks, vol. 946, pp. 271–283. Springer, Singapore (2023). https://doi.org/10.1007/978-981-19-5868-7_21
7. Campanella, G., Silva, V.W.K., Fuchs, T.J.: Terabyte-scale deep multiple instance learning for classification and localization in pathology. *arXiv preprint arXiv:1805.06983* (2018)
8. Cao, R., et al.: Prostate cancer detection and segmentation in multi-parametric MRI via CNN and conditional random field. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1900–1904. IEEE (2019). <https://ieeexplore.ieee.org/abstract/document/8759584/>
9. Carneiro Neto, P.: Deep learning based analysis of prostate cancer from MP-MRI, Master's thesis (2020). <https://aaltodoc.aalto.fi/handle/123456789/44951>
10. Chun, C.W., Yousir, N.T., Abdulameer, S.M., Mostafa, S.A., Hezam, A.A.: Deep learning approach for predicting prostate cancer from MRI images. *J. Soft Comput. Data Min.* **3**(2), 1–9 (2022). <https://publisher.uthm.edu.my/ojs/index.php/jscdm/article/view/12220>

11. De Vente, C., Vos, P., Hosseinzadeh, M., Pluim, J., Veta, M.: Deep learning regression for prostate cancer detection and grading in bi-parametric MRI. *IEEE Trans. Biomed. Eng.* **68**(2), 374–383 (2020). <https://ieeexplore.ieee.org/abstract/document/9090311/>
12. Gülmez, B.: A novel deep neural network model based Xception and genetic algorithm for detection of COVID-19 from X-ray images. *Ann. Oper. Res.* **328**(1), 617–641 (2023)
13. Hassan, E., Shams, M.Y., Hikail, N.A., Elmougy, S.: A novel convolutional neural network model for malaria cell images classification. *Comput. Mater. Continua* **72**(3), 5889–5907 (2022). <https://doi.org/10.32604/cmc.2022.025629>
14. Janiesch, C., Zschech, P., Heinrich, K.: Machine learning and deep learning. *Electron. Mark.* **31**(3), 685–695 (2021)
15. Kavitha, R., et al.: Retracted: ant colony optimization-enabled CNN deep learning technique for accurate detection of cervical cancer. *Biomed. Res. Int.* **2023**(1), 1742891 (2023)
16. Kundu, N.: Exploring ResNet50: an in-depth look at the model architecture and code implementation (2023). <https://medium.com/@nitishkundu1993/exploring-resnet50-an-in-depth-look-at-the-model-architecture-and-code-implementation-d8d8fa67e46f>
17. Li, W., et al.: Path R-CNN for prostate cancer diagnosis and Gleason grading of histological images. *IEEE Trans. Med. Imag.* **38**(4), 945–954 (2018). <https://ieeexplore.ieee.org/abstract/document/8490855/>
18. Litjens, G., et al.: Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci. Rep.* **6**(1), 26286 (2016). <https://www.nature.com/articles/srep26286>
19. Liu, S., Zheng, H., Feng, Y., Li, W.: Prostate cancer diagnosis using deep learning with 3D multiparametric MRI. In: *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134, pp. 581–584. SPIE (2017). <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10134/1013428/Prostate-cancerdiagnosis-using-deep-learning-with-3D-multiparametric-MRI/10.1117/12.2277121.short>
20. Marks, L.S., et al.: PCA3 molecular urine assay for prostate cancer in men undergoing repeat biopsy. *Urology* **69**(3), 532–535 (2007). <https://www.sciencedirect.com/science/article/pii/S0090429506026380>
21. Mun, Y., Paik, I., Shin, S.J., Kwak, T.Y., Chang, H.: Yet another automated Gleason grading system (YAAGGS) by weakly supervised deep learning. *NPJ Digital Med.* **4**(1), 99 (2021). <https://www.nature.com/articles/s41746-021-00469-6>
22. Nagpal, K., et al.: Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *NPJ Digital Med.* **2**(1), 48 (2019). <https://www.nature.com/articles/s41746-019-0112-2>
23. Naik, N., et al.: Role of deep learning in prostate cancer management: past, present and future based on a comprehensive literature review. *J. Clin. Med.* **11**(13), 3575 (2022). <https://www.mdpi.com/2077-0383/11/13/3575>
24. Nakasi, R., Mwebaze, E., Zawedde, A., Tusubira, J., Akera, B., Maiga, G.: A new approach for microscopic diagnosis of malaria parasites in thick blood smears using pre-trained deep learning models. *SN Appl. Sci.* **2**(7), 1255 (2020)
25. Neto, P.C.: Deep learning based analysis of prostate cancer from MP-MRI. arXiv preprint [arXiv:2106.01835](https://arxiv.org/abs/2106.01835) (2021)
26. Patel, A., Singh, S.K., Khamparia, A.: Detection of prostate cancer using deep learning framework. In: *IOP Conference Series: Materials Science and Engineering*, vol. 1022, p. 012073. IOP Publishing (2021). <https://iopscience.iop.org/article/10.1088/1757-899X/1022/1/012073/meta>

27. Provenzano, D., et al.: Machine learning algorithm accuracy using single-versus multi-institutional image data in the classification of prostate MRI lesions. *Appl. Sci.* **13**(2), 1088 (2023). <https://www.mdpi.com/2076-3417/13/2/1088>
28. Rostami, B., Anisuzzaman, D.M., Wang, C., Gopalakrishnan, S., Niezgoda, J., Yu, Z.: Multiclass wound image classification using an ensemble deep CNN-based classifier. *Comput. Biol. Med.* **134**, 104536 (2021). <https://www.sciencedirect.com/science/article/pii/S0010482521003309>
29. Ryu, H.S., et al.: Automated Gleason scoring and tumor quantification in prostate core needle biopsy images using deep neural networks and its comparison with pathologist-based assessment. *Cancers* **11**(12), 1860 (2019). <https://www.mdpi.com/2072-6694/11/12/1860>
30. Salman, M.E., Cakar, G.C., Azimjonov, J., Kosem, M., Cedimoglu, I.H.: Automated prostate cancer grading and diagnosis system using deep learning-based Yolo object detection algorithm. *Expert Syst. Appl.* **201**, 117148 (2022). <https://www.sciencedirect.com/science/article/pii/S0957417422005425>
31. Singh, D., Vishwakarma, M., Sharma, A., Singh, S.: Detection of prostate cancer using recurrent neural network (2024). <https://www.irejournals.com/formatedpaper/1705472.pdf>
32. Society, A.C.: Key statistics for prostate cancer, prostate cancer facts (2024). <https://www.cancer.org/cancer/types/prostate-cancer/about/key-statistics.html>
33. Ström, P., et al.: Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol.* **21**(2), 222–232 (2020). [https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045\(19\)30738-7/fulltext?13570](https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(19)30738-7/fulltext?13570)
34. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826 (2016)
35. Talaat, F.M., El-Sappagh, S., Alnowaiser, K., Hassan, E.: Improved prostate cancer diagnosis using a modified ResNet50-based deep learning architecture. *BMC Med. Inf. Decis. Mak.* **24**(1), 23 (2024)
36. Tsehay, Y., et al.: Biopsy-guided learning with deep convolutional neural networks for Prostate Cancer detection on multiparametric MRI. In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 642–645. IEEE (2017). <https://ieeexplore.ieee.org/abstract/document/7950602/>
37. Veeling, B.S., Linmans, J., Winkens, J., Cohen, T., Welling, M.: Rotation equivariant CNNs for digital pathology. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) *MICCAI 2018*. LNCS, vol. 11071, pp. 210–218. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_24
38. Wang, X., et al.: Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning. *Sci. Rep.* **7**(1), 15415 (2017). <https://www.nature.com/articles/s41598-017-15720-y>
39. Xu, D., et al.: Mask R-CNN assisted 2.5 D object detection pipeline of 68Ga-PSMA-11 PET/CT-positive metastatic pelvic lymph node after radical prostatectomy from solely CT imaging. *Sci. rep.* **13**(1), 1696 (2023). <https://www.nature.com/articles/s41598-023-28669-y>



Empowering Healthcare Queries: A Chatbot for Converting Text to SQL

Nyashadzashe Tamuka¹(✉)  and Khulumani Sibanda² 

¹ University of Fort Hare, Alice, Eastern Cape, South Africa
201516429@ufh.ac.za

² Walter Sisulu University, East London, Eastern Cape, South Africa
ksibanda@wsu.ac.za

Abstract. In the healthcare industry, retrieving accurate patient and medical data in real time is essential for decision-making. However, healthcare providers often lack the technical expertise to query databases. Many chatbots struggle with ambiguous questions from non-SQL users, leading to inaccurate or incomplete SQL queries. Additionally, existing models lack error correction and alternative query formulation. This study presents a text-to-SQL chatbot for healthcare professionals to access medical data using natural language. We compared ‘ChatGPT-4o’ and ‘all-MiniLM-L6-v2’ for handling ambiguous healthcare queries and error correction. Microsoft Azure was used to develop the healthcare database, integrating Named Entity Recognition and context management for improved efficiency. Results showed that ChatGPT-4o outperformed all-MiniLM-L6-v2, achieving 92% execution accuracy and 90% error-handling success, highlighting the chatbot’s effectiveness in assisting non-technical users.

Keywords: Named Entity Recognition (NER) · Context Management · Chatbot · text-to-SQL · ChatGPT-4o · all-MiniLM-L6-v2 · Healthcare queries · Error-handling · Natural Language Processing (NLP)

1 Introduction

Healthcare organizations generate massive volumes of data such as electronic health records and patient treatment histories. Efficient data retrieval from these databases is crucial for healthcare providers to make informed clinical decisions. Traditionally, SQL has been widely adopted for querying relational databases, yet many healthcare professionals lack the technical expertise to write SQL queries. Consequently, essential patient information may be underutilized, leading to delays in care or incomplete assessments. AI assistants (chatbots), specifically text-to-SQL systems, provide a solution by enabling users to retrieve records using their language. These systems translate human language into SQL queries that can interact with a database, thus eliminating the need for specialized knowledge in writing queries.

The establishment of large language models (LLMs) such as GPT and sentence transformers has significantly advanced the field of natural language processing (NLP),

making it possible to design chatbots that can accurately interpret user queries and generate valid SQL commands. While many sectors such as e-commerce and finance have benefited from chatbots, the healthcare industry has been slower to adopt such technologies due to the accuracy of the generated queries. This study aimed to improve the performance of text-to-SQL chatbots, particularly for the healthcare industry by handling ambiguous questions and providing error correction follow-up responses.

2 Background

2.1 Related Studies

Several studies have presented the adoption of chatbots (AI assistants) in the healthcare sector. The studies highlight various approaches to improve chatbot functionality.

A related study by [1] proposed a machine learning-based chatbot that utilized a deep learning approach to answer patients' frequently asked questions (FAQs). This convolutional neural network (CNN) model classified queries into predefined categories and responded accordingly. The chatbot performed well in simple interactions, but its limitations were exposed when handling more complex medical inquiries. Despite its ability to generate responses for general questions, the study did not indicate the system's ability to generate SQL commands for ambiguous queries.

Similarly, [2] proposed a reinforcement learning (RL) chatbot for mental health care. This chatbot was designed to interpret patients' sentiments such as stress and provide medical procedures. By integrating RL, the chatbot could improve its decision-making capabilities through user feedback, which led to more accurate responses over time. However, the system faced challenges in interpreting ambiguous or incomplete queries, as well as generating precise SQL queries for medical data retrieval.

A study by [3] explored the use of transformer-based models, such as BERT, to design a chatbot capable of interpreting patient queries and retrieving clinical data. This study demonstrated that transformer models significantly improved query understanding compared to traditional machine-learning approaches. Nevertheless, the chatbot's ability to convert complex, natural language queries into accurate SQL queries for healthcare databases was not fully explored, leaving a gap in its practical application.

A study by [4] proposed the use of GPT to develop conversational agents for healthcare data retrieval. The study indicated that GPT-based assistants can be applied to interact with healthcare providers and patients, interpreting natural language inputs and generating responses. The study showed that although GPT-based assistants outperform traditional rule-based chatbots in query understanding, they have some limitations such as data privacy and data interpretation. While GPT's language model enhanced conversation, the study did not address challenges such as query disambiguation and the correction of invalid SQL commands.

A related study by [5] proposed an AI-based chatbot for patient record management. It was developed using the treatment records and symptoms data collected from the DoctorMe application. The study demonstrated an improvement in query interpretation accuracy, particularly when handling medical terminology. However, the system's ability to generate SQL queries for real-time data retrieval was still limited. The study did

not indicate the chatbot's performance with ambiguous queries and error correction mechanisms were not highlighted.

Despite advancements in chatbot technology for healthcare, the existing studies have not fully addressed the issue of generating valid SQL queries from ambiguous natural language inputs, particularly for non-technical users. Moreover, most studies have focused on improving conversational abilities without integrating error correction mechanisms for invalid queries. Motivated by these limitations, the aim of this study was to apply Named Entity Recognition (NER) and context management to develop a chatbot that provides query disambiguation and follow-up error correction.

3 Methods

3.1 Model Architecture for Text-To-SQL Chatbot

The proposed chatbot model architecture was designed to translate user natural language queries into SQL queries for efficient interaction with the relational *Victoria* database. Microsoft Azure, which is a cloud platform was used to host our database. To process ambiguous queries and provide corrections to invalid queries, we utilized a large language model, ChatGPT-4o [6]. The model's API token which was used to develop the proposed health care assistant was accessed from our OpenAI profile. ChatGPT-4o was compared with an 'all-MiniLM-L6-v2' model which is a sentence transformer, known for providing optimal accuracy and speed for generating SQL queries [7].

Database Schema

The chatbot was connected to the *Victoria* database which consisted of the *Patients* table and the *MedicalRecords* table. The dataset used was not real, but it was synthesized using ChatGPT for reflecting the common characteristics from the electronic health care dataset such as patient demographics and billing information. The synthetic dataset consisted of features from real world electronic health datasets such as patient demographics and billing information. The *Patients Table* which contained 2000 rows stored patient information, including patient_id, first_name, last_name, date_of_birth, contact_number, and email. Figure 1 shows the Patient's table schema.

The *MedicalRecords Table* which consisted of 2000 rows, stored medical information for each patient, including patient_id, diagnosis, treatment_plan, medications, and billing_amount. Figure 2 indicates the MedicalRecords table's schema.

Python programming was used to develop the chatbot's backend and React Native was utilized for establishing the frontend. Figure 3 indicates the chatbot's architecture.

The following sections present the Chatbot's components:

3.2 User Query Handling

The chatbot system began with the UserQuery class, which processed the user's natural language. The user query handling consisted of:

Context Management - The system retrieved the user's query history to maintain conversational flow. Using the Context class, the chatbot stored and retrieved the context of past interactions, allowing it to better refine future queries.

```

CREATE TABLE Patients (
    patient_id INT PRIMARY KEY,           -- Unique ID for each patient
    first_name VARCHAR(100),              -- Patient's first name
    last_name VARCHAR(100),               -- Patient's last name
    gender VARCHAR(10),                   -- Gender of the patient (Male, Female, Other)
    date_of_birth DATE,                   -- Patient's date of birth
    contact_number VARCHAR(15),           -- Contact number for the patient
    email VARCHAR(255),                   -- Email address for the patient
    address VARCHAR(255),                 -- Patient's residential address
    city VARCHAR(100),                    -- City of residence
    state VARCHAR(100),                   -- State of residence
    zip_code VARCHAR(10),                 -- ZIP code of the patient's area
    insurance_provider VARCHAR(100),      -- Insurance provider information
    insurance_number VARCHAR(100),        -- Patient's insurance policy number
    registration_date DATE                -- Date of patient registration
);

```

Fig. 1. The schema for the Patient's table.

```

CREATE TABLE MedicalRecords (
    record_id INT PRIMARY KEY,             -- Unique ID for each medical record
    patient_id INT,                        -- Foreign key referencing the Patients table
    visit_date DATE,                       -- Date of the patient visit
    doctor_id INT,                         -- ID of the doctor who attended the patient
    diagnosis VARCHAR(255),                -- Diagnosis details of the patient
    treatment_plan TEXT,                   -- Detailed treatment plan
    prescription TEXT,                     -- Prescribed medications
    follow_up_date DATE,                   -- Follow-up date for the patient
    total_cost DECIMAL(10, 2),              -- Total cost of the medical services
    payment_status VARCHAR(50),             -- Payment status (Paid, Unpaid, Pending)
    FOREIGN KEY (patient_id) REFERENCES Patients(patient_id) -- Relationship to the Patients table
);

```

Fig. 2. The schema for the *MedicalRecords* table.

Query Normalization - Before further processing, unnecessary keywords such as “this”, “the”, “a” and “an” were removed from the query to simplify its structure. This step was crucial for improving the accuracy of the natural language to SQL conversion.

3.3 Named Entity Recognition (NER)

Named Entity Recognition is the information extraction task in Natural Language Processing that involves identifying and classifying entities such as a person, date, organization, animal, and other entities [8]. The NER class was applied to handle queries from non-technical users who may not know exact column names. For instance, users

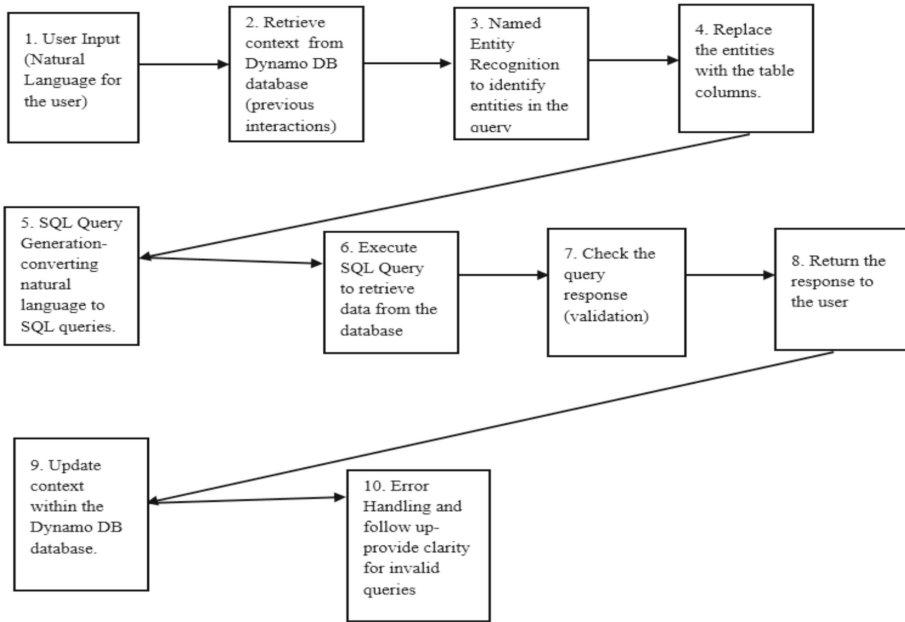


Fig. 3. Chatbot’s architecture.

may refer to “patients” or “records” instead of Patients or MedicalRecords tables. NER was responsible for mapping these terms to the actual SQL components. In the context of the dataset we used, entities such as patients’ names and dates were identified and replaced by the corresponding table field names such as first_name, and date_of_birth. The chatbot system used the NER class to detect keywords such as “insurance” and “cost” and mapped them to tables’ fields such as “insurance_provider” and “total_cost”. This was an essential stage to disambiguate the user’s natural language.

Data Augmentation

Data augmentation refers to enriching the input data to improve the chatbot’s ability to generate accurate and relevant SQL queries [9]. This involved developing a *Table-Model* class that included the tables’ schemas. We provided the chatbot with a detailed description of the tables and fields to provide more context using the *TableModel* class. This was applied to disambiguate vague queries from the users. For example, a user can query “Show me Voster’s age?” This query can be vague if there is a patient with a first name Voster and another patient with the last name Voster or if there are multiple patients with the first name, Voster. By applying data augmentation, we provided details about the tables and their fields. The chatbot could identify Voster as a first name or last name based on the *TableModel* class. Figure 4 shows the *TableModel* class for data augmentation.

```

table_example_data="""
patient_id,first_name,last_name,gender,date_of_birth,contact_number,email,address,city,state,zip_code,insurance_provider,insurance_num
1,John,Doe,Male,1985-06-15,+1234567890,john.doe@example.com,123 Main St,,Sandton,10001,ProviderX,12345,2022-01-01
2,Jane,Smith,Female,1998-08-21,+1987654321,jane.smith@example.com,456 Elm St,San Capetown,CA,94107,ProviderY,67890,2021-05-15
3,Michael,Johnson,Male,1975-12-03,+1122334455,michael.johnson@example.com,789 Oak St,East London,CA,90001,ProviderZ,11223,2020-12-30
4,Sarah,Davis,Female,1988-09-25,+2233445566,sarah.davis@example.com,101 Pine St,Johannesburg,FL,33101,ProviderX,44556,2021-10-10
"""

# Defining the MedicalRecords Table schema and example data
medical_records_table = TableModel(
    table_description="""The MedicalRecords table stores each patient's medical history, diagnoses, treatments,
    prescribed medications, and billing information.""",
    table_schema="""
CREATE TABLE MedicalRecords (

```

Fig. 4. *TableModel* class for data augmentation.

3.4 Query Execution

After generating the SQL query, the chatbot executed it against the *Victoria* database. The SQL query was executed using the *execute_sql_query* method from the *QueryGenerator* class. We integrated a retry method that handled execution errors, such as invalid SQL commands. If a query failed, the chatbot asked follow-up questions to refine the query and guide the user toward a correct result.

3.5 Context Management

The Context class stored the user's interaction history in a DynamoDB database so the chatbot could handle multi-turn conversations (multiple conversations between the user and the chatbot). When a user submits a follow-up query, the system could refer to the stored context to refine the query, for instance, questions that apply joins between the tables. For example, a user could first ask, "Show me patient Ivan records," and then follow up with, "What was his diagnosis?" The context management system will remember that "Ivan" refers to a patient's name in the *Patients* table, and "diagnosis" is a field in the *MedicalRecords* table, allowing it to generate a query joining the two tables.

3.6 Error Detection and Query Refinement

If an SQL query produced incorrect results, the *is_response_incorrect* method checked the results based on expected columns and data types. The chatbot provided follow-up questions to the user for query refinement by asking for more details.

Evaluation Metrics

Several metrics such as execution accuracy, error handling success rate, logical accuracy, conversational ability, and response time were applied for the evaluation of the chatbot's performance.

Execution Accuracy

Evaluates whether the SQL query established by a chatbot returned the correct result when executed against the database. It can be expressed as (1):

$$\text{Execution accuracy} = \frac{\text{Number of correctly executed queries}}{\text{Total number of queries}} * 100\% \quad (1)$$

Logical Accuracy

Shows the ability of a chatbot to generate an SQL query that matches the expected one, even if the query's execution might not return correct results. It can be shown as (2):

$$\text{Logical accuracy} = \frac{\text{Number of correctly structure queries}}{\text{Total number of queries}} * 100\% \quad (2)$$

Error Handling Success Rate

Shows the ability of the chatbot to process ambiguous queries by either correcting them or prompting the user for clarification. It can be shown as (3):

$$\text{Error handling success rate} = \frac{\text{Number of correctly handled ambiguous queries}}{\text{Total number of ambiguous queries}} * 100\% \quad (3)$$

Conversational Ability

Shows the ability of the chatbot to maintain context over multi-turn conversations, which is important for refining queries or handling follow-up questions. It can be shown as (4):

$$\text{Conversational ability} = \frac{\text{Number of correctly maintained conversations}}{\text{Total number of multi – turn conservartion}} * 100\% \quad (4)$$

Response Time

The amount of time the chatbot generates a SQL query from the user's natural language.

4 Results

This section presents the findings of developing and testing the text-to-SQL chatbot (assistant) for retrieval of healthcare data using natural language queries. The chatbot's performance was evaluated using metrics which are execution accuracy, logical form accuracy, error handling success rate, conversational ability, and response time. The results for two models, the ChatGPT-4o and all-MiniLM-L6-v2 are compared.

4.1 Execution Accuracy

The chatbot was tested using 25 different queries which included straightforward queries like "Show me all patient records" and more complex and ambiguous queries like "What is Xiang's diagnosis?" where multi-table joins, and disambiguation were required. ChatGPT-4o achieved an execution accuracy of 92%, correctly interpreting 23 out of 25 queries. The all-MiniLM-L6-v2 had an execution accuracy of 88% but struggled with complex queries that included joins.

4.2 Logical Accuracy

ChatGPT-4o yielded a logical accuracy of 84%, showing high proficiency in generating SQL queries with appropriate syntax. Out of 25 queries, ChatGPT-4o generated 21 SQL queries that matched the intended structure (ground truth). For example, the chatbot was asked “*Show me all the patients from East London?*” The generated query was:

```
SELECT * FROM Patients WHERE city = 'East London';
```

The chatbot was asked, “*Show me the medical records of Xiang Robertsons*”. The generated query was:

```
SELECT * FROM MedicalRecords  
WHERE patient_id = (SELECT patient_id FROM Patients WHERE  
first_name = ' Xiang ' AND last_name = ' Robertsons');
```

The all-MiniLM-L6-v2 achieved a logical accuracy of 80%. The all-MiniLM-L6-v2 failed to generate correct queries for 5 natural language questions such as “*How much is Xiang Robertsons treatment*” and “*What was Sarah’s latest visit to the hospital?*”.

4.3 Error Handling Success Rate

ChatGPT-4o achieved a 90% error-handling success rate. The chatbot using the ChatGPT-4o model prompted the user for additional information or provided follow-up questions for 9 out of 10 ambiguous questions such as:

“*What is Sarah’s age*”

“*Thus incorrect*”

“*Show me her insurance details*”

“*What is the patient’s registration date?*”

On the other hand, the chatbot using the all-MiniLM-L6-v2 had an error-handling success rate of 50%. It struggled with 5 ambiguous questions and hallucinated (the model provided wrong responses).

4.4 Conversational Ability

The ChatGPT-4o exhibited the best conversational capacity. The ChatGPT-4o model retained context across 95% of multi-turn conversations. For example, the ChatGPT-4o model was asked a question: “*What was Sarah Smith’s latest visit to the hospital?*”. Its response was 20/02/2020. The follow-up question was “*Show me her diagnosis*”. The ChatGPT-4o’s response was Covid-19. This showed the model’s superior performance as opposed to the all-MiniLM-L6-v2 model which exhibited 75% of multi-turn interactions. The all-MiniLM-L6-v2 model failed to retrieve context for ambiguous questions. For example, a question: “*What was Sarah Smith’s latest visit to the hospital?*”. The all-MiniLM-L6-v2 model provided a correct response which is 20/02/2020. However, the all-MiniLM-L6-v2 model struggled with a follow-up question: “*Show me her diagnosis*”. Its response was “*Please provide more details*”, which is a generic response that we developed for out-of-context questions.

4.5 Response Time

The ChatGPT-4o model had a response time of 1.8 s per query. The all-MiniLM-L6-v2 outperformed ChatGPT-4o in terms of speed, with an average response time of 1.5 s per query. This indicates that the all-MiniLM-L6-v2 sentence transformer was faster in generating SQL queries from text. Table 1 shows the comparison of the two models we evaluated for developing a healthcare assistant (chatbot).

Table 1. Comparison of the ChatGPT-4o and the all-MiniLM-L6-v2 models.

Metric	ChatGPT-4o	all-MiniLM-L6-v2
Execution Accuracy	92%	88%
Logical Accuracy	84%	80%
Error Handling Success Rate	90%	50%
Conversational Ability	95%	75%
Response Time (seconds)	1.8s	1.5s

5 Discussion

The ChatGPT-4o model demonstrated superior capabilities in handling complex SQL queries, especially those requiring table joins and subqueries. Its ability to maintain context across multi-turn interactions was important for refining queries. The high logical accuracy suggests that ChatGPT-4o generated SQL queries that aligned closely with the intended (actual). While the all-MiniLM-L6-v2 was faster in terms of response time, its accuracy in generating valid SQL queries was slightly lower, particularly for complex and vague queries. This model performed well in straightforward queries but required more frequent error corrections in multi-turn conversations. Both models indicated a higher success rate in generating accurate queries with correct joins, but the all-MiniLM-L6-v2 had a higher misinterpretation rate for nested queries requiring multiple conditions as compared to the ChatGPT-4o model. The application of chatbots (assistants) in the healthcare sector has gained substantial attention. However, handling ambiguous queries and providing follow-up error correction responses is a black box. The proposed text-to-SQL chatbot system outperformed various related chatbot models by providing effective query disambiguation and error correction mechanisms. In comparison to the CNN-based chatbot proposed by [1], which struggled with complex medical inquiries and could not generate accurate SQL queries, our chatbot not only processed complex queries but also generated accurate SQL commands. Our ChatGPT-4o-based chatbot offered follow-up questions for ambiguous user queries as opposed to the reinforcement learning assistant proposed by [2] which could not handle ambiguous user queries. Moreover, the transformer-based chatbot developed by [3] excelled in understanding patient queries but did not fully explore the ability to convert these queries into SQL commands. In

contrast, our chatbot incorporated Named Entity Recognition (NER) and context management for SQL query generation and processing of vague queries. Although the NER model significantly improved query accuracy, the ChatGPT-4o-based chatbot yielded a higher success rate in interpreting ambiguous medical terms as compared to the all-MiniLM-L6-v2 model. The GPT chatbot proposed by [4] improved conversational flow in the healthcare sector. Nevertheless, the study did not address query disambiguation or correction of out-of-context queries, limitations that were improved by our proposed chatbot. Furthermore, the patient record management chatbot by [5] handled medical terminology well but lacked SQL generation and error correction, which are strengths of our system.

6 Conclusion

This study presents the development and evaluation of a text-to-SQL chatbot designed for the healthcare sector. By utilizing models such as ChatGPT-4o and the all-MiniLM-L6-v2 sentence transformer, the chatbot successfully translates natural language queries into SQL. ChatGPT-4o showed higher execution and logical form accuracy, particularly in handling complex queries, while the all-MiniLM-L6-v2 offered faster response times for simpler queries. The chatbot's ability to handle ambiguous queries and maintain multi-turn conversations demonstrated its potential for application in the healthcare sector. Part of future work includes integrating voice to SQL functionality and a variety of natural languages to accommodate diversity in the healthcare sector. Our proposed chatbot which uses the Named Entity Recognition (NER) and context management, not only benefits the healthcare sector but can be applied in other sectors such as education, banking, transport, and precise agriculture.

References

1. Rai, S., Raut, A., Savaliya, A., Shankarmani, R.: March. Darwin: convolutional neural network-based intelligent health assistant. In: 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 1367–1371. IEEE (2018)
2. Sohara Banu, A.R., Nagaveni, V., Gowda, N.C., Yerriswamy, T., Gururaj, M., Varshitha, N.: Well bot—a reinforcement learning and sentiment analysis-based chatbot implementation for mental well-being. In: Reddy, V.S., Wang, J., Reddy, K. (eds.) *Soft Computing and Signal Processing*. ICSCSP 2023. LNNS, vol. 864. Springer, Singapore (2023). https://doi.org/10.1007/978-981-99-8628-6_40
3. Babu, A., Boddu, S.B.: BERT-based medical chatbot: enhancing healthcare communication through natural language understanding. *Explor. Res. Clin. Soc. Pharm.* **13**, 100419 (2024)
4. Javaid, M., Haleem, A., Singh, R.P.: ChatGPT for healthcare services: An emerging stage for an innovative perspective. *BenchCouncil Trans. Benchmarks Stand. Evaluat.* **3**(1), 100105 (2023)
5. Dammavalam, S.R., Chandana, N., Rao, T.R., Lahari, A., Aparna, B.: AI-based chatbot for the hospital management system. In: 2022 3rd International Conference on Computing, Analytics and Networks (ICAN), pp. 1–5. IEEE (2022)
6. OpenAI. (2023). <https://openai.com/>. Accessed 19 Sept 2024
7. Jain, S.M.: Hugging face. In: *Introduction to transformers for NLP: With the hugging face library and models to solve problems*, pp. 51–67. Apress, Berkeley, CA (2022)

8. Li, J., Sun, A., Han, J., Li, C.: A survey on deep learning for named entity recognition. *IEEE Trans. Knowl. Data Eng.* **34**(1), 50–70 (2020)
9. Shorten, C., Khoshgoftaar, T.M., Furht, B.: Text data augmentation for deep learning. *J. Big Data* **8**(1), 101 (2021)



HEALeB: Healthcare 4.0 Enabled by IoT, AI, and Blockchain

Henriette Kyangwa¹(✉) , Landry Mbale¹ , and Antoine Bagula^{1,2}

¹ Université Nouveaux Horizons, Lubumbashi, Democratic Republic of the Congo
henriengonga@gmail.com

² University of the Western Cape, Cape Town, South Africa
abagula@uwc.ac.za

Abstract. As the volume and sensitivity of medical data continue to grow, ensuring data security and effective analysis has become increasingly challenging. This paper presents the design and development of HEALeB, a comprehensive system that integrates blockchain technology and machine learning (ML) to address these challenges. Blockchain technology offers a decentralized, immutable ledger that ensures data integrity, privacy, and traceability, while ML algorithms enhance predictive analytics and decision-making capabilities. By combining these technologies, HEALeB aims to overcome critical issues such as data tampering, unauthorized access, and the underutilization of decision-support tools. The system's design leverages blockchain to secure Electronic Health Records (EHRs) and employs ML to provide advanced analytics for personalized and precise healthcare solutions. This approach promises to improve diagnostic accuracy, optimize health management, and enhance overall healthcare quality. The paper provides a detailed exploration of HEALeB's architecture, its implementation, and its performance evaluation, concluding with insights into the system's impact and future research directions in the realm of Healthcare 4.0.

Keywords: Blockchain · Machine Learning · Medical Data Security · Predictive Analytics · Healthcare Data Management

1 Introduction

In the era of Healthcare 4.0, the healthcare sector is experiencing a rapid transformation driven by the integration of advanced digital technologies such as the Internet of Things (IoT), Artificial Intelligence (AI), and Blockchain. With over two zettabytes of medical data generated annually, largely due to the widespread adoption of mobile health applications—now exceeding 350,000 globally—there is immense potential for improving healthcare outcomes through data-driven insights and automation. However, this exponential growth in data introduces several critical challenges, particularly in terms of data management, security, and analysis.

One of the most pressing challenges in Healthcare 4.0 is ensuring the confidentiality, integrity, and traceability of Electronic Health Records (EHRs). Securing sensitive health information while providing authorized healthcare providers with timely access to data is crucial for effective patient care. Furthermore, the underutilization of decision-support tools in many medical environments can hinder diagnostic accuracy and delay critical decision-making processes. Another significant challenge lies in the development of predictive diagnostic strategies to anticipate health risks, particularly in regions with limited healthcare infrastructure. Overcoming these challenges requires the deployment of sophisticated technological solutions that can secure data and enhance healthcare analytics in a seamless, efficient manner.

1.1 Blockchain Technology for Medical Data Security

Blockchain technology has emerged as a pivotal component of Healthcare 4.0, offering a decentralized and immutable ledger system that enhances the security of medical data. By decentralizing data storage, blockchain ensures that sensitive health information is protected from unauthorized access or tampering. Additionally, blockchain creates a transparent platform for storing and sharing medical records, allowing healthcare professionals to access real-time patient information and make more informed decisions. This technology is also revolutionizing the pharmaceutical supply chain, ensuring the authenticity of medications and tracking each step of their journey from production to delivery. Furthermore, in clinical trials, blockchain facilitates greater transparency by creating immutable records of research protocols and results, reducing the risk of fraud and improving the accessibility of trial data for future research.

1.2 Machine Learning for Data Analysis

Machine Learning (ML) plays an integral role in Healthcare 4.0 by enabling the advanced analysis of large and complex datasets. ML algorithms can identify patterns and trends in patient data, predict potential health outcomes, and suggest personalized treatment plans. Through predictive analytics, ML can forecast health events such as disease outbreaks or patient deterioration, allowing healthcare providers to implement preventive measures that improve patient outcomes. Moreover, ML facilitates personalized medicine by tailoring treatments to individual patients based on their genetic information, lifestyle, and medical history. In the management of EHRs, ML algorithms enhance early disease detection, optimize treatment plans, and streamline administrative processes like appointment scheduling, thereby improving both diagnostic accuracy and overall care efficiency.

1.3 Challenges, Contributions, and Outline

Healthcare 4.0 introduces both opportunities and challenges, particularly in the realms of medical data security and analytics. The integration of blockchain and

machine learning represents a powerful approach to addressing these challenges. Blockchain technology ensures that healthcare data is secure, immutable, and accessible only by authorized individuals, while machine learning leverages this data to extract valuable insights that improve decision-making and enable personalized care. Together, these technologies form a robust framework for addressing the challenges faced by the healthcare sector in managing, securing, and utilizing medical data effectively.

This paper presents the design and development of **HEALeB**, an integrated system that combines blockchain for secure medical data management with machine learning for advanced data analytics. HEALeB addresses key challenges related to data security, privacy, and the underutilization of decision-support tools in healthcare environments. By improving diagnostic precision, optimizing treatment plans, and enabling personalized healthcare solutions, HEALeB aligns with the broader goals of Healthcare 4.0.

The remainder of this paper is organized as follows: Sect. 2 reviews the relevant literature on blockchain and machine learning in healthcare. Section 3 introduces blockchain technology and its applications in medical data security. Section 3.1 explores the role of machine learning in healthcare data analysis. Section 3.2 presents the architecture and design of the **HEALeB** system. Section 4 evaluates the system's performance, and Sect. 5 concludes with insights into challenges and future research directions within the context of Healthcare 4.0.

2 Literature Review

This section covers two key areas in healthcare technology: the use of machine learning (ML) for disease prediction and the integration of blockchain for securing healthcare data, particularly in predictive healthcare models. We explore the significant contributions in each field and discuss how both technologies can complement each other to enhance the quality and security of healthcare systems.

2.1 Machine Learning for Disease Prediction

Machine learning techniques have been widely applied in healthcare for disease prediction, especially in cardiovascular diseases. Table 1 summarizes several studies that have utilized different machine learning algorithms to predict heart diseases. The table provides a comparative analysis of various techniques and their performance. The studies in Table 1 highlight the progress in disease prediction, particularly heart disease, using a variety of machine learning models such as SVM, Decision Trees, and Random Forests. Many of these techniques achieve high accuracy rates, demonstrating the effectiveness of machine learning in improving healthcare diagnostics.

Table 1. Machine Learning Techniques Applied in Healthcare for Disease Prediction

Year	Autors	Techniques	Accuracy	Observation
2010	Son et al. [13]	SVM	77.63%	Machine learning algorithm Support Vector Machine (SVM) for heart failure prediction
2016	Saqlain et al. [10]	DT, LR, RF, ANN, NB, SVM	86.6% (NB)	Identification of Heart Failure by Using Unstructured Data of Cardiac Patients
2018	Polat et al. [7]	Adaboost, DT, LR, RF, SVM	87.1% (LR)	Application of machine learning for disease prediction
2019	Ali et al. [1]	SVM linear, SVM RBF	92.22% (SVM)	An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure
2020	Deepika et Sasikala [4]	DT	83.6%	Enhanced Model for Prediction and Classification of Cardiovascular Disease using Decision Tree with Particle Swarm Optimization
2021	Rubini et al. [9]	RFC, LR, SVM, NB	84.80% (RFC), 83.82% (LR)	Cardiovascular Disease Prediction using Machine Learning Algorithms
2022	Gupta et al. [5]	KNN, LR, SVM, DT, RF, SVM	92.3% (LR)	Heart disease prediction using supervised ML techniques
2023	Bhatt et al. [3]	DT	72.77%	Effective Heart Disease Prediction Using Machine Learning Techniques

2.2 Blockchain for Data Security in Healthcare

While machine learning improves predictive accuracy, the integration of blockchain technology is essential for securing the sensitive healthcare data that feeds into these predictive models. Blockchain offers solutions to the key issues of data integrity, privacy, and compliance with regulations such as HIPAA and GDPR.

Blockchain is particularly relevant in managing **Electronic Health Records (EHRs)**, where it provides decentralized and tamper-proof data storage. The immutable nature of blockchain ensures that data once recorded cannot be altered, thereby maintaining data integrity. Additionally, advanced cryptographic techniques in blockchain protect sensitive health information and ensure that only authorized parties can access the data.

Recent advancements in blockchain applications in healthcare, as summarized by the article [2, 8], show its significant role in managing medical records, remote patient monitoring, and securing the supply chain of pharmaceutical products. The distribution of blockchain use cases in the healthcare sector, as noted by Singh et al. [12], highlights the growing adoption of this technology for various purposes:

- 11% in remote patient monitoring
- 48% in medical data recording
- 21% in data analysis and other uses
- 15% in biomedical research
- 5% in drug supply chain management

2.3 The Intersection of Machine Learning and Blockchain

Machine learning models require large datasets for accurate predictions. The use of IoT devices in healthcare has dramatically increased the volume of data collected, especially in areas like remote patient monitoring. This influx of data needs to be securely managed, which is where blockchain comes in. Blockchain provides the security framework necessary for protecting medical data collected through IoT devices and used by machine learning models. The article [11] presents a study proposing a multimodal approach, utilizing BERT-CNN for standardization and game theory to optimize performance, while enhancing security through homomorphic encryption and the Ethereum blockchain. The integration of these technologies improves the secure, scalable, and decentralized management of healthcare data. With a 23% increase in accuracy, this method proves to be more effective than current approaches for medical data classification. Our studies aims to further improve these results by employing additional techniques and focusing specifically on cardiac applications. As noted by Patel et al. [6], combining machine learning with blockchain can improve both predictive accuracy and data security. Blockchain secures the data used by machine learning algorithms, ensuring privacy and integrity, while machine learning provides valuable insights by analyzing this secure data. The combination of these technologies is critical for future healthcare systems, particularly in areas that require high levels of data security and compliance with privacy regulations such as HIPAA, the POPI Act, and GDPR.

3 HEALeB: Integration of Blockchain and Machine Learning in Healthcare 4.0

The **HEALeB** system represents an advanced integration of blockchain and machine learning technologies within healthcare systems, aimed at addressing critical issues related to medical data security, privacy, and analysis. By leveraging the strengths of both technologies, HealeB provides a comprehensive solution for managing patient data, enhancing diagnostic accuracy, and supporting personalized care. Some of its key features include:

Data Security and Privacy: At the core of the HealeB system is blockchain technology, which provides a secure and decentralized framework for managing sensitive medical data. This ensures that patient information remains protected from tampering or unauthorized access. The immutability of blockchain further guarantees data integrity, while smart contracts enable automated, secure access to medical records by authorized users. This decentralized control empowers patients to manage who can access their data, increasing transparency and trust in the healthcare system.

Enhanced Analytical Accuracy: HealeB leverages machine learning algorithms to analyze the securely stored data. Blockchain's reliability as a source of data ensures that ML models work with accurate, tamper-proof information, leading to better predictive analytics and improved diagnostic precision.

The combination of these technologies enhances personalized treatment plans by analyzing patient data in real time, supporting healthcare professionals in making data-driven decisions for improved patient outcomes.

Compliance and Auditability: The transparency and traceability offered by blockchain are particularly useful for regulatory compliance. HealeB’s immutable audit trails ensure that all data access and modifications are securely recorded, making it easier for healthcare organizations to comply with industry regulations. Additionally, blockchain smart contracts ensure that healthcare transactions adhere to predefined rules, automating processes while maintaining auditability.

3.1 HEALeB Deployment Scenario

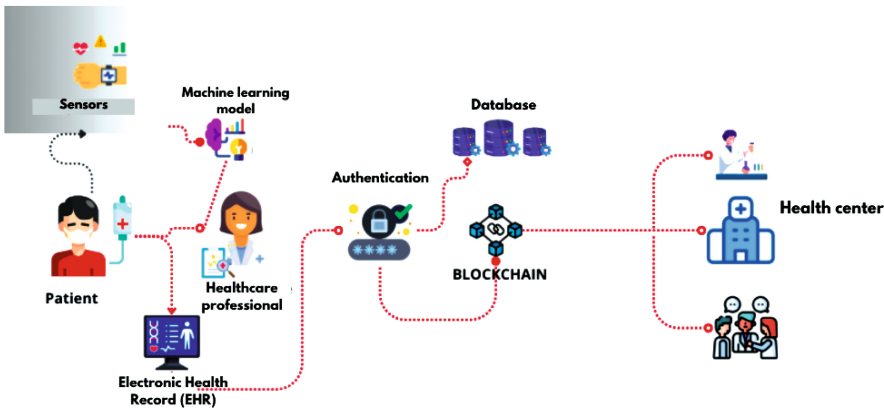


Fig. 1. HealBe Deployment Scenario

Figure 1 illustrates a deployment scenario in which blockchain and machine learning technologies collaborate to manage and analyze patient data. The scenario involves several key steps. First, sensors continuously monitor and gather data on patients’ vital signs, including heart rate, blood pressure, and oxygen levels. This data is then transmitted to a machine learning server, where it undergoes analysis to generate insights. These insights may involve pattern recognition, predicting health risks, or suggesting possible treatments. Health professionals subsequently review and validate the results to ensure their accuracy and relevance.

After analysis, the data is authenticated to confirm its integrity and accuracy before it is securely recorded. The data is logged into a database and added to the blockchain, where it is stored immutably. This immutability protects the data from tampering or unauthorized modifications. Finally, authorized users such as healthcare providers and researchers can securely access the data through applications. The transparency provided by the blockchain allows users to trace

the data's history and verify its authenticity, enhancing trust and reliability in the information used for patient care.

This deployment scenario integrates human oversight, particularly during the data analysis phase, ensuring alignment with emerging regulations and policies that emphasize the importance of human judgment in medical data processing (Fig. 2).

3.2 HEALeB System Architecture

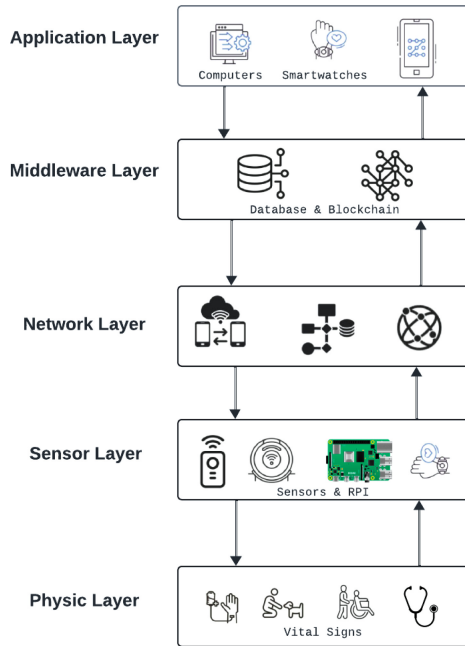


Fig. 2. The HealeB system architecture, showing the interaction between the physical, sensor, network, middleware, and application layers.

The architecture of **HEALeB** consists of multiple layers, each representing a different functional component of the system. The layers are structured to ensure effective collection, transmission, and processing of health data, ultimately providing secure, actionable insights to users. The layers are described below:

1. Physical Layer (Couche Physique): The physical layer forms the foundation of the system, where vital signs are collected. This layer represents physical interactions with patients, such as monitoring heart rate, blood pressure, and other physiological metrics. Medical devices used in hospitals or wearable technologies are employed to gather this essential health data.

2. Sensor Layer (Couche de Capteurs): The sensor layer is responsible for gathering data through sensors and devices. These include wearables, medical IoT devices, and Raspberry Pi systems. The devices collect real-time information from the patient and convert it into digital form, enabling further processing by the system.

3. Network Layer (Couche Réseau): This layer ensures the secure and efficient transmission of the health data collected by the sensor layer. It comprises various network protocols and Internet infrastructure that facilitate communication between the sensors and higher-level components. Reliable transmission mechanisms are crucial for sending data to storage and processing units.

4. Middleware Layer (Couche Middleware): The middleware layer acts as an intermediary between the network and application layers. It includes:

- **Database:** A system that securely stores the collected health data for future retrieval and analysis.
- **Blockchain:** This ensures the integrity and security of sensitive medical data. Blockchain provides a decentralized ledger for storing patient data, ensuring traceability and protection from unauthorized tampering.

5. Application Layer (Couche Applicative): The application layer is the user-facing component of the system. It includes interfaces and applications running on devices such as:

- **Smartphones:** Mobile applications for patients and healthcare providers to access, visualize, and interact with health data.
- **Computers:** Desktop applications used in hospitals and clinics to access health records and diagnostic tools.

This layer provides actionable insights from the data collected and processed in the lower layers, facilitating real-time healthcare management.

4 Performance Evaluation and Results

The **HEALeB** system was evaluated in a real-world healthcare setting to assess its performance in terms of data security, privacy, and analytics capabilities. The results demonstrated that the integration of blockchain technology and machine learning significantly improved the accuracy and efficiency of medical data management and analysis. The system successfully addressed key challenges, such as unauthorized access, data tampering, and underutilization of decision-support tools, thereby enhancing the overall quality of healthcare delivery.

The results were obtained by applying various classification algorithms to the dataset. We employed five statistical learning algorithms and conducted a comparative analysis. The highest accuracy rate was observed with the KNN, SVM, LR, DT, and RF algorithms, as demonstrated by the results depicted in Fig. 3. The Random Forest algorithm proved to be an effective approach for the classification of cardiovascular diseases. It demonstrated a high accuracy of

	Classifier	Accuracy	ROC_AUC	Recall	Precision	F1
4	Random Forest	91.670000	0.950000	0.890000	0.950000	0.920000
3	Decision Tree	88.400000	0.920000	0.860000	0.910000	0.890000
1	Nearest Neighbors	84.620000	0.890000	0.780000	0.910000	0.840000
0	Logistic Regression	71.150000	0.750000	0.630000	0.780000	0.700000
2	Support Vectors	53.150000	0.560000	0.390000	0.580000	0.470000

Fig. 3. Model performances of various algorithms.

91.67%, as illustrated in Fig. 3, with the ability to correctly classify a significant percentage of test samples.

We further extended our analysis by examining the relationship between the false positive rate and the true positive rate using the ROC curve (Fig. 4).

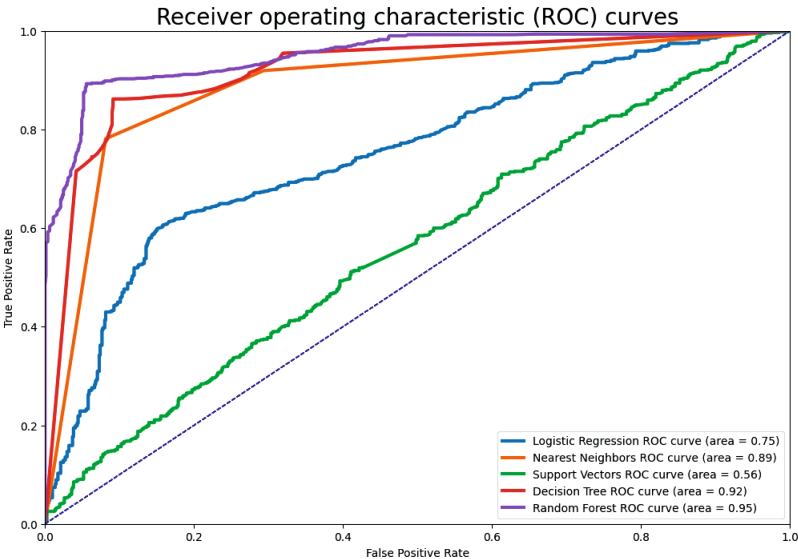


Fig. 4. ROC curve analysis of the models.

By utilizing modern machine learning algorithms, specifically boosted trees like CatBoost, XGBoost, and LightGBM, which are optimized machine learning algorithms based on the gradient boosting technique, we found that XGBoost achieved the best performance with an accuracy of 88.87% (Fig. 5).

The confusion matrices (CMs) for various classification algorithms are presented below. Each matrix visually represents the classification results for Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), CatBoost, XGBoost, and LightGBM models. The matrices are displayed three per line for efficient space utilization (Figs. 6, 7, 8, 9, 10, 11, 12 and 13).

	Classifier	Accuracy	ROC_AUC	Recall	Precision	F1
1	xgboost	88.870000	0.940000	0.850000	0.930000	0.890000
2	light GBM	86.890000	0.930000	0.830000	0.910000	0.870000
0	Catboost	86.480000	0.930000	0.830000	0.910000	0.870000

Fig. 5. Performance comparison of boosted trees algorithms.

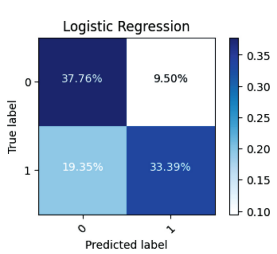


Fig. 6. LR CM

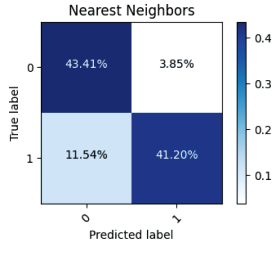


Fig. 7. KNN CM

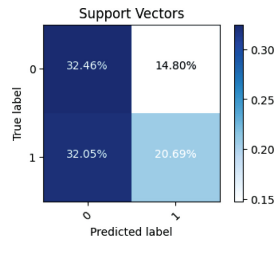


Fig. 8. SVM CM

To further improve the performance, we applied hyperparameter tuning on the XGBoost model. By focusing on key parameters such as ‘num-leaves’, ‘max-depth’, and ‘min-data-in-leaf’, we achieved an increase in accuracy from 88% to 93%. Moreover, we observed a reduction in the false positive rate from 5.1% to 2.4%, and the false negative rate from 5.5% to 4.8% (Figs. 14 and 15).

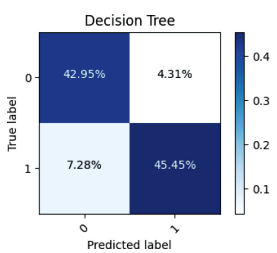


Fig. 9. DT CM

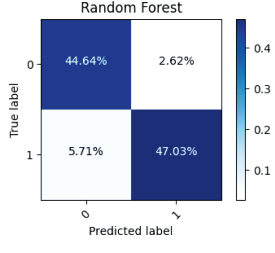


Fig. 10. RF CM

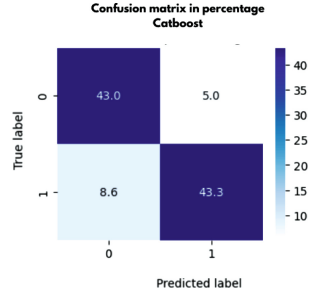


Fig. 11. CatBoost CM

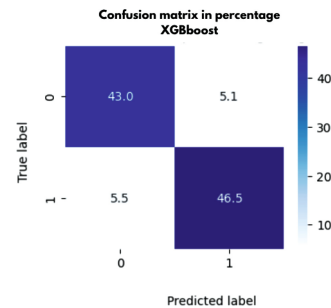


Fig. 12. XGBoost CM

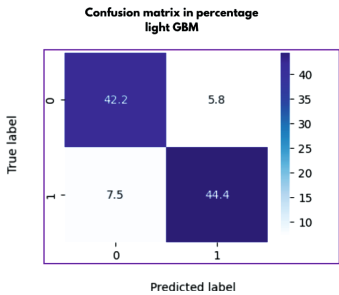


Fig. 13. LightGBM CM

	precision	recall	f1-score	support
0	0.90	0.95	0.93	811
1	0.95	0.91	0.93	905
accuracy			0.93	1716
macro avg	0.93	0.93	0.93	1716
weighted avg	0.93	0.93	0.93	1716

Fig. 14. Performance of XGBoost after hyperparameter tuning.

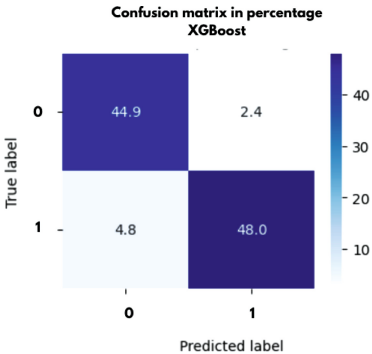


Fig. 15. Confusion Matrix of XGBoost after tuning.

5 Conclusion and Future Research Directions

The **HEALeB** system integrates blockchain and machine learning to address critical challenges in medical data security, privacy, and analytics. By combining blockchain’s decentralized, tamper-proof data management with machine learning’s predictive capabilities, **HEALeB** enhances diagnostic precision, optimizes treatment plans, and ensures compliance with regulations like HIPAA and GDPR. The system improves healthcare delivery by providing secure and personalized care while protecting sensitive patient information.

Performance evaluations have shown that **HEALeB** significantly improves the security and accuracy of medical data management. Machine learning models used for disease prediction achieved high accuracy, while blockchain ensured the integrity and privacy of patient data.

Future research should focus on improving scalability, interoperability with existing EHR platforms, and testing in diverse healthcare environments. Integrating edge computing and IoMT could further enhance real-time data analysis and patient monitoring. Additionally, exploring advanced machine learning models and privacy-preserving techniques like homomorphic encryption could improve predictive accuracy and data security.

In summary, **HEALeB** demonstrates a significant step toward Healthcare 4.0 by combining secure data management and advanced analytics. Ongoing research into scalability, real-time capabilities, and privacy will ensure that **HEALeB** continues to evolve and meet the needs of modern healthcare systems.

References

1. Ali, L., et al.: An optimized stacked support vector machines based expert system for the effective prediction of heart failure. *IEEE Access* **7**, 54007–54014 (2019)
2. Bathula, A., et al.: Blockchain, artificial intelligence, and healthcare: the tripod of future - a narrative review. *Artif. Intell. Rev.* **57**, 238 (2024)

3. Bhatt, C.M., Patel, P., Ghetia, T., Mazzeo, P.L.: Effective heart disease prediction using machine learning techniques. *Algorithms* **16**(2), 88 (2023)
4. Deepika, P., Sasikala, S.: Enhanced model for prediction and classification of cardiovascular disease using decision tree with particle swarm optimization. In: 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 1068–1072. IEEE (2020)
5. Gupta, C., Saha, A., Reddy, N.S., Acharya, U.D.: Cardiac disease prediction using supervised machine learning techniques. In: *Journal of Physics: Conference Series*, vol. 2161, p. 012013. IOP Publishing (2022)
6. Patel, R., Dwivedi, V.: A review of secure IoT based smart health monitoring system using blockchain technique. In: 2024 3rd International Conference on Sentiment Analysis and Deep Learning (ICSADL), pp. 569–580. IEEE (2024)
7. Polat, K., Şahan, S., Güneş, S.: Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k-nn (nearest neighbour) based weighting preprocessing. *Expert Syst. Appl.* **32**(2), 625–631 (2007)
8. Ramzan, S., Aqduş, A., Ravi, V., Koundal, D., Amin, R., Al Ghamdi, M.A.: Healthcare applications using blockchain technology: motivations and challenges. *IEEE Trans. Eng. Manag.* **70**(8), 2874–2890 (2023)
9. Rubini, P., Subasini, C., Katharine, A.V., Kumaresan, V., Kumar, S.G., Nithya, T.: A cardiovascular disease prediction using machine learning algorithms. *Ann. Rom. Soc. Cell Biol.* **25**, 904–912 (2021)
10. Saqlain, M., Hussain, W., Saqib, N.A., Khan, M.A.: Identification of heart failure by using unstructured data of cardiac patients. In: 2016 45th International Conference on Parallel Processing Workshops (ICPPW), pp. 426–431. IEEE (2016)
11. Shingari, N., Mago, B.: The importance of data encryption in ensuring the confidentiality and security of financial records of medical health. In: 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), vol. 2, pp. 1–6 (2024)
12. Singh, Y., Jabbar, M.A., Kumar Shandilya, S., Vovk, O., Hnatiuk, Y.: Exploring applications of blockchain in healthcare: road map and future directions. *Front. Public Health* **11**, 1229386 (2023)
13. Son, Y.-J., Kim, H.-G., Kim, E.-H., Choi, S., Lee, S.-K.: Application of support vector machine for prediction of medication adherence in heart failure patients. *Healthc. Inform. Res.* **16**(4), 253–259 (2010)

Smart Systems Enabling Technologies



Comparative Analysis of AI-Based Algorithms for the Detection of Tuberculosis Bacilli on Digitised Human Tissue

Kapongo D. Lumamba^{1,2(✉)}, Threnesan Naidoo^{2,4}, Mpumelelo Msimang⁵,
Gordon Wells², Adrie J. C. Steyn^{2,3}, and Mandlenkosi V. Gwetu^{1,6}

¹ School of Mathematics, Statistics and Computer Science, University of Kwazulu Natal (UKZN), King Edward Avenue, Scottsville, Pietermaritzburg 3209, KwaZulu Natal, Republic of South Africa

lumamkd@gmail.com, gwetum@ukzn.ac.za

² Africa Health Research Institute, UKZN, 719 Umbilo Road, Durban 10587, KwaZulu Natal, Republic of South Africa

{threnesan.aidoo, adrie.steyn}@ahri.org, msimangm@ukzn.ac.za

³ Department of Microbiology, University of Alabama, 1720 2Nd Ave South, Birmingham, AL 35294, USA

⁴ Department of Forensic and Legal Medicine, Walter Sisulu University, Nelson Mandela Dr, Umtata Part 1, Mthatha 5099, Eastern Cape, Republic of South Africa

⁵ Department of Anatomical Pathology, National Health Laboratory Service, Inkosi Albert Luthuli Central Hospital, Durban, South Africa
mpumelelo.msimang@nhls.ac.za

⁶ Department of Industrial Engineering, University of Stellenbosch, Stellenbosch 7600, South Africa

Abstract. Computer vision techniques, particularly those employing Artificial Intelligence (AI) and Machine Learning (ML), have shown promise in the analysis of tuberculosis (TB) using digitized human tissue specimens. Our systematic review identified two prominent studies that applied deep learning frameworks to detect mycobacteria in digital pathology slides. Despite their reported successes in other contexts, our collaborative efforts to replicate and extend these models yielded unexpected results, prompting a thorough investigation into the potential causes behind their underperformance. This study comparatively analyses the deployment of two prominent models to the task of image-based TB detection. The research reveals that key factors contributing to underperformance include dataset limitations, model architecture complexities, annotation quality issues, transferability constraints, preprocessing discrepancies, and clinical relevance variations. While deep learning computer vision models hold significant potential for enhancing TB diagnosis through the analysis of human tissue specimens, our collaborative testing of two prominent models revealed several challenges that may hinder their performance in real-world applications. Addressing these challenges will be crucial for advancing the reliability and effectiveness of such models in clinical practice.

Keywords: Deep Learning · Image Processing · Mycobacterium · Tuberculosis · Human Tissue

1 Introduction

Tuberculosis (TB) is a that can be traced back approximately 9,000 years as per findings from archaeological remains in Atlit Yam, city off Israel's coast. It affects the human lungs primarily and was first identified by Robert Koch in 1882 [5, 10]. "One would think that the pathology of tuberculosis would have been accurately described long ago and that science in the 21st century would have advanced far beyond morphologic descriptions" [12], but such is not the case. The *Mycobacterium tuberculosis* (*Mtb*) bacteria mainly attack alveoli, causing inflammation and damage to lung tissue which is still a challenge even with advancements as per WHO reports of over 1.30million deaths globally each year due to *Mtb* [27] despite science's progress since Koch's era [5, 7]. The human lung is a complicated structure found in the thoracic cavity and plays a critical function in breathing and gas exchange. Various structural components form parts of the human lung, including the trachea, bronchi, bronchioles, and alveoli – the site of gas exchange [16]. The pathological assessment of human lung matters can offer helpful information concerning the progress of *Mtb* disease and help towards treatment decisions.

Simultaneously, the world witnessed considerable growth in multimedia use on the internet during the early 2000s which catalysed computer vision evolution from a research area into an important technology widely accepted and used today [4]. The initial breakthrough of AlexNet Network by its inventors set this trajectory [13]. At the heart of computer vision is image processing, manipulating static images or video sequences to extract valuable information essential for diagnosis and treatment in medical fields like TB detection where understanding lung anatomy plays a crucial role [10, 28, 30]. This discipline has roots dating back at least as early as 1970s when techniques were developed using X-ray based tomography, offering three-dimensional body imaging [3, 6].

Further advancements in digital imagery and sophisticated image processing algorithms came into play during the 80's and 90's [9, 26]. Computer Aided Diagnosis systems emerged to assist medical professionals by identifying patterns indicative of diseases using computer-generated analysis [30]. Recently, machine learning has made significant strides in imaging analyses with special mention towards deep learning algorithms offering tremendous potential for tumour detection and treatment. As technology continues its relentless march forward, it is expected that image processing will continue to gain prominence especially given the vast amounts of data available today [22, 30].

Deep learning (DL) is an instance of the machine learning paradigm of feature learning [19, 22]. Iteratively, it expands upon acquired representations of the primary data with class separability attainment goal [6, 19]. In the last few years, DL has had an immense impact on numerous disciplines of technology. Its latest improvements facilitated the identification, classification, and quantification of patterns in medical imaging [7]. Machine learning techniques preceding deep learning era in Computer Vision field include Decision Trees – a model based on partitioning or splitting a dataset

into partitions [6, 20]. Artificial Neural Network first conceptualised in 1940's and became practical with the development of computing machinery which mimic the human brain's neurons to identify patterns [14]. ANNs are like DL in that they involve layers of nodes, but they typically have far fewer layers than deep neural networks [14, 31]. Support Vector Machines are used specifically for problems pertaining to classification [23] whereas Nearest Neighbours involves locating data points that are more similar [21, 25]. Bagging and Boosting algorithms [11] were also some of the commonly utilized techniques, though Deep Learning has quickly become the primary approach due to its capability of managing complex unstructured datasets [23, 26]. In this section a general introduction of the topic is described. In the next section we provide a brief description of two machine learning models in the context of TB.

2 Summary of Studies

The field of computer vision has indeed made significant strides in the study and diagnosis of cancer compared to tuberculosis (TB). Several factors contribute to this discrepancy. Our previous literature review research (18) identified 5 prominent studies that applied CNN-based frameworks to detect mycobacteria in digital pathology slides as an attempt to reduce the TB diagnostic time for Pathologists and improve investigative correctness by developing machine learning models as either a support to the Pathologist or a standalone investigative tool. The following 2 studies out of the 5 were willing to collaborate: "A CNN-based active learning framework to identify mycobacteria in digitized Ziehl-Neelsen (ZN) stained human tissues, 2020" (hereafter referred as **Model A**), and "Deep-Learning-Aided Detection of *Mycobacteria* in Pathology Specimens Increases the Sensitivity in Early Diagnosis of Pulmonary Tuberculosis Compared with Bacteriology Tests, 2022" (referred to as **Model B**). The following subsections briefly introduce the two reviewed studies, emphasising their areas of focus and reported findings.

2.1 Study 1 – Model A

In 2020, [30] proposed an Artificial Intelligence assisted diagnostic system using a combination of a Convolutional Neural Network model (Inception-V3) for patch-based classification and a Logistic Regression (LR) model for the classification of the Whole Slide Image (WSI). From the input WSIs, the model first extracted patches (256×256 pixels). The extracted patches were then used in the CNN model training, which consisted of detecting Acid-Fast Bacilli (AFB) on the patches. Consequently, the model was retrained using a semi-supervised Active Learning (AL) framework on additional AFB negative patches from 19 WSIs. Then patches were classified as either "positive" or "negative" using the retrained CNN model. The LR model then used the results from the CNN classification to label the entire slide as either AFB "positive" or "negative". Superimposed onto the inputted WSI, was a score heat map generated by the model to indicate the likelihood of AFB. Diagnosticians would then inspect each region within the heat maps to confirm the results. The study used the Aperio scanner, AT version (Leica Biosystems) to digitise 14 glass slides into WSIs but the authors did not clearly declare the patch labelling platform used (Fig. 1).

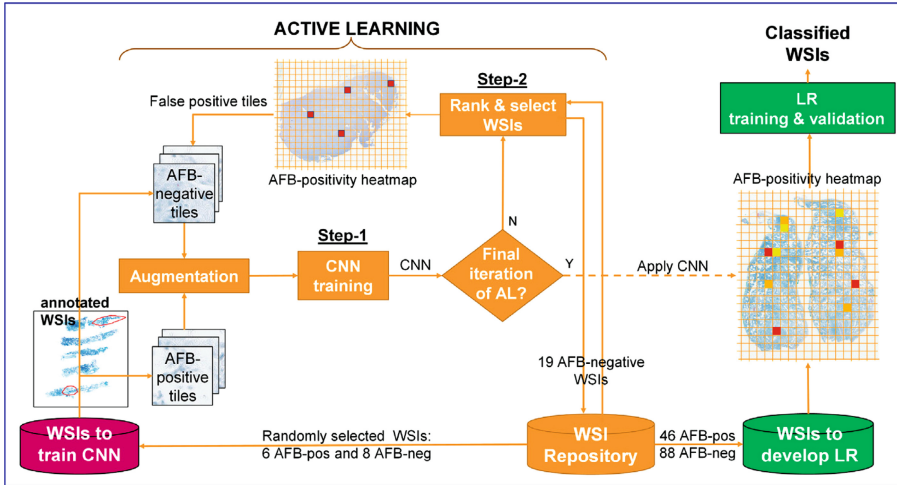


Fig. 1. System overview [30].

The patches frameworks from this model [30] were validated using a diverse set of patches and accomplished 99.03% and 98.75% for F1 scores. The performance metrics at the WSI level were 87.13%, 87.62%, 80.18% for the sensitivity, specificity, and F1 respectively for the entire pipeline. Overall, a mean average precision (mAP) of up to 96.41% was achieved.

2.2 Study 2 – Model B

Zaisen et al. [32] established an algorithm for detecting AFBs in ZN-stained human tissue sections using a pre-trained HALO AI CNN, referred to as an “AI-supported diagnosing method”. Each patch identified as potentially positive by the algorithm went through consensus evaluation by six diagnosticians.

The researchers introduced an advanced deep learning techniques for detecting the presence of *Mycobacterium tuberculosis* in human pathology specimens, which they established to be superior when it comes to early diagnosis as compared with standard bacteriological tests. This model was trained and checked using 1,123 samples from patients suspected of pulmonary tuberculosis (PTB), yielding a sensitivity of 94% along with a specificity rate of 92%. Furthermore, the outcome of a comparison of results between the model and bacteriology tests was evident that the sensitivity of the DL method was extensive in the PTB early detection. To accomplish their task, they procured samples from two autopsy cases of pulmonary TB and 40 instances of biopsy wherein AFBs were present but not initially detected via conventional methods for training the model, using Motic EasyScan (Motic, Hong Kong, China) to digitise glass slides while employing HALO software as a tool in labelling and instruction phase. In their practical application involving clinical tests with 42 cases including patients whose mycobacteriosis was confirmed by bacteriological techniques or developed during follow-ups, the deep learning assisted method achieved commendable performance stats of sensitivity at 86% alongside full specificity (100%).

Despite the exceptional results, there are some limitations worth noting; like the study used WSI generated from only one slide scanner and as AFB positive training data, only autopsy cases of TB were used. Another probable limitation is that the suggested method might wrongly identify other bacterial species as *Mycobacterium*. In contrast, the method from **Study 1** [30] engaged pathologists to assess a score heatmap overlaying the analysed WSI, to facilitate the assessment of suspected regions in the context of histopathological structures. Nonetheless, **Study 2** [32] did not provide a specific report on how the platform was employed for analysis by Pathologists. Instead, the authors mentioned that six Pathologists were responsible for examining each potentially positive patch for confirmation. Additionally, the models depend on the quality of the input pathology images. Unreliable outcomes are likely to be produced if the input samples were of reduced quality or had artifacts for various reasons.

3 WSI Datasets

The TB whole-slide image datasets are collections of digitized and annotated human tissue samples. These samples serve as a valued resource for academics and medical professionals to develop and test algorithms that can accurately and effectively detect TB pathogens in histopathological images. The development of these datasets is an important step towards improving the diagnosis' accuracy and speed. Traditionally, the diagnosis of TB has relied on a combination of clinical symptoms and lab tests, which can be time-consuming and can delay treatment. Researchers can make use of the TB WSI datasets to, for instance, train machine learning algorithms to accurately identify TB pathogens in histopathological images. In the future, the same can be used for the development of an accurate and targeted treatment for TB. From the already existing TB WSI datasets, the following are some of the most used:

- Radbound University Medical Center (RUMC) dataset which consists of 18,000 WSI of TB and non-TB lung tissue samples from 1,000 patients. It is one of the largest TB WSI datasets available [29].
- Warwick University dataset – contains 863 WSI of TB and non-TB lung tissue samples. It is annotated with detailed region-level labels to facilitate the development of more accurate TB detection algorithms [22].
- Zhejiang University dataset - consisting of 327 WSI of TB and non-TB lung tissue samples, as well as annotations for nuclei and mitotic figures. It is a valuable resource for developing a more complex machine learning algorithm [33].
- Institute of Cancer Research (ICR) dataset – concludes 500 WSI of TB and non-TB lung samples, as well as annotations for epithelial cells, lymphocytes, macrophages, and other important structures [17].

While these TB WSI datasets represent a valuable resource for the development of more accurate TB diagnosis algorithms, there are some limitations or shortfalls that should be considered. The following are some of the notable ones:

- Limited diversity [15] – Much TB WSI datasets suffer from a lack of demographic diversity, particularly in terms of race and ethnicity. This can lead to bias in the

development of TB diagnosis algorithms, which may not be effective for patients from more diverse backgrounds.

- Limited sample sizes – while some TB WSI datasets are quite large, others may have relatively small sample sizes, which can limit their usefulness for the improvement of robust and accurate TB diagnosis algorithms [1].
- Limited range of TB subtypes – most TB WSI datasets focus on a specific subtype of the disease, such as pulmonary TB. This may limit their usefulness for the diagnosis and treatment of other subtypes [8].
- Annotated data vs. real-world data – while many TB WSI datasets are annotated with detailed labels and metadata, this may not precisely exhibit the complexity of real-world TB diagnosis scenarios [2]. This can make it difficult to generalize findings from TB WSI datasets to real-world clinical settings.
- Lack of standardization – there is currently a lack of standardization across different TB WSI datasets in terms of imaging protocols, and data annotations criteria [24]. This can make it difficult to compare results across different studies and datasets.

In summary, though the TB WSI datasets are a valuable resource for the development of more accurate TB diagnosis algorithms, it is important to be aware of their limitations to ensure that their findings are effectively translated into a real-world clinical setting.

4 WSI Datasets

TB WSI datasets are an important resource for improving the accuracy and speed of TB diagnosis and have the potential to significantly improve the detection and treatment of the disease. TB imaging datasets may be smaller, less standardized, and more challenging to obtain due to issues such as privacy concerns, variability in image acquisition protocols, and access to human tissue. We are positioned in a very unique space that grants us access to a variety cohort of human TB tissue samples through various collaborators. Based on the limitations of existing datasets, a more comprehensive collection is being assembled. The purpose of this study is to explore the effectiveness of existing algorithms in this new collection, in order to identify opportunities for contextualising TB detection algorithms. All scripts and data necessary to reproduce the analyses are available in the public repository hosted on GitHub: <https://github.com/lumamkd/Comparative-Analysis-of-AI-based-Algorithms>.

4.1 Data Collection

While attempting to develop a larger and more diverse TB WSI dataset of over 1000 WSI from multiple human organ tissue, we have selected a small sample set of 24 WSI from digitized human tissue samples as depicted below (Table 1) for the preliminary comparison. The selected samples came from 13 different patients (6 females, 6 males, and 1 not disclosed) from the Department of Anatomical Pathology, National Health Laboratory Service, Inkosi Albert Luthuli Central Hospital (IALCH), Durban, South Africa. The surgical resection was conducted in two different hospitals – IALCH and King Dinizulu Hospital between 2007 and 2020. The formalin fixed waxed blocks generated from these samples were cut and ZN-stained in different periods between 2013 to 2023. All the 24 glass slides were digitalised in 2024.

4.2 Preprocessing of Tissue Samples

The selected samples consist of 23 human lung tissue samples and 1 testicular tissue sample. All samples are from surgical resection procedure. The tissues were formalin fixed from the date of surgery (Table 1), then later dissected for cassettes and tissue processed using the Tissue-Tek VIP 6 AI Vacuum Infiltration Processor machine from Sakura. Following the xylene free processing of the tissue, they were then wax embedded and stained with H&E and ZN.

Table 1. Selected dataset with AFB status from Pathology Diagnosis

SID	Yr Surgery	AFB	Age	Sex	Organ
S01	2007				LUNG
S02					LUNG
S03					LUNG
S04					LUNG
S05	2013	++	35	M	L LUNG
S06		+	32	F	L LUNG
S07		+	46	M	L LUNG
S08		+++	32	M	R LUNG
S09		–	32	M	LUNG
S10		–		F	R LUNG
S11	2018		50	M	R LUNG
S12					
S13					
S14					
S15					
S16					
S17					
S18	2018		33	F	R LUNG
S19					
S20					
S21			57	F	L LUNG
S22					
S23	2020	+++	38	M	L Testis
S24					

4.3 Preprocessing of Tissue Samples

The ZN-stained glass slides were digitalised twice, first using the Motic EasyScan 1.0.5.13.4 (Motic, Hong Kong, China) that produced the *.svs format WSI. Then we used the Hamamatsu NanoZoomer S60 (C13210) for the *.ndpi format. One WSI (S24) was received from the authors of Model B, and 23 tiles from Model A.

4.4 Equipment

All applications are installed on virtual machine running Windows 10 Education with an Intel® Xeron® Gold 6226 CPU @ 2.70GHz, 64GB of RAM and accessed via VMware Horizon Client (version 2306, build 8.10.0). The machine is assigned 16 core of the CPU and has NVIDIA Quadro RTX 8000 attached. Files (WSIs, tiles and csv) are stored in a separate file server. For Model A, we used the Jupyter Notebook 7.0.8 under Anaconda, while HALO v3.6.4134.396 and HALO AI 3.6.4134 were used for Model B testing.

5 Machine Learning Models for AFB Detection – In House Testing

There is significant potential for the use of the proposed machine learning methods in medical finding and disease management, since they offer a precise, fast, and effective [30] means of checking and detecting *mycobacteria* in digitised human tissue samples, allowing timely detection and treatment. With potential to better sensitivity and specificity, the method by [32] could have considerable impact on public health programs targeted at stopping the spreading of tuberculosis pathogens.

5.1 Model A

Description. The Authors [30] conducted a retrospective study in the United States of America (USA) that was published in 2020. To identify *mycobacteria* in digitised ZN-stained human tissue, the study employed a CNN based active learning framework. This framework combined the power of CNN based deep learning with active learning, which involves selecting informative samples for annotation to improve model performance.

After signing a user agreement, we received a compressed folder containing the model file (*.h5), python script (Script #1), a requirement text file and a sample set of 23 images (tiles from WSIs) that were used on the model training. Later, the authors provided us with an additional python script (Script #2) that splits the WSI into tiles after we have written our own (Script #3) with main differences in colour normalization and size down sampling. To recreate a similar setting where the model was trained and validated, we had to follow the exact steps and requirement parameters. Unfortunately, the authors could not share the entire programme as published in their article. The main parameters were to set the image resolution to 0.5 μ /px and the tiles size to 256×256 pixels (Fig. 2).

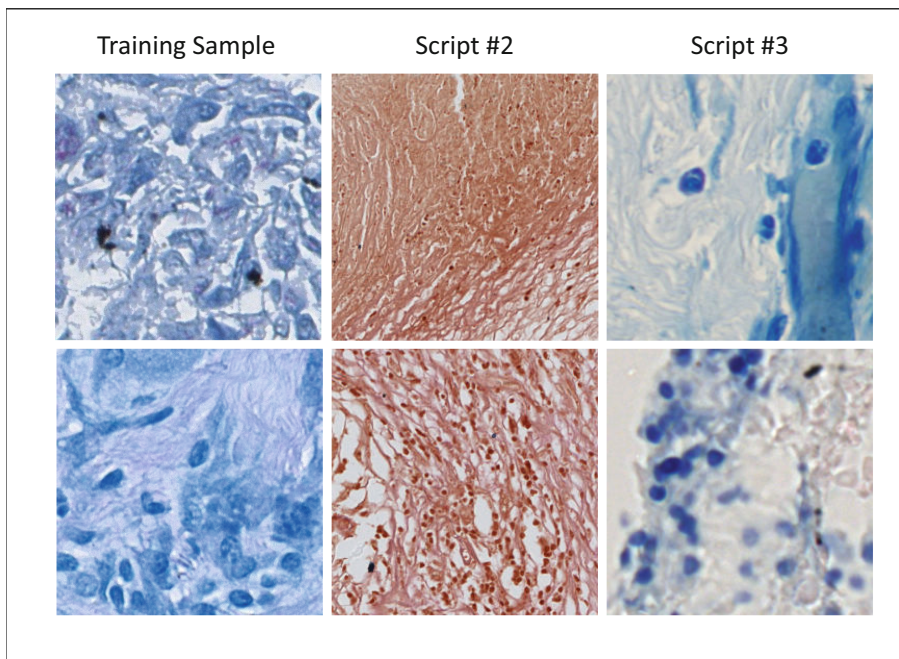


Fig. 2. Sample of image tiles generated from different python scripts.

For each of the imaging format (*.svs and *.ndpi), we adapted scripts #2 and #3 to read the WSI, split it into tiles and set the resolution of each tile to the required value before saving them on separate folders. Refer to Table 2 for the total number of tiles generated. It was observed that setting the recommended image resolution on the WSI before splitting drained the machine memory and caused the process to crash.

Initially, script #1 would simply point to the folder containing sample tiles, analyse them and display the result of each tile on the jupyter notebook. Script #1 was then modified to access the main folder and loop through a number of sub-folders created by scripts #2 and #3 that represent each WSI. The modified script #1 is hereafter referred to as script #1.0 which creates a *.csv file for each sub-folder using the WSI base name. Then reads the tiles and analyses them using the model as received from the authors and appends the results in the *.csv file. Using the *.csv file, the last script #4 collected all the positive tiles names into a list, then used the list to move the positive tiles of each WSI into a new sub-folder for review. From Table 2, it is important to note the huge difference in the number of generated tiles between scripts #2 and #3 from both image modalities.

Performance and Limitations. First, we assessed script #1 using the training sample tiles shared by the authors and the results were as expected. Second, we tested the model using script #1.0 on tiles generated by scripts #2 and #3 from both WSI formats (*.svs and *.ndpi) separately and the results are shown on Table 2. We observe that for every WSI, there are positive tiles meaning tiles that contain at least one AFB regardless of

which script was used for splitting. However, on close examination, majority of the tiles predicted as positive do not contain an AFB.

Time limitation – based on our computational settings, it would take over 24 h to complete the pipeline from accessing one WSI from the file server, splitting into tiles and analysing. When the WSI and the tiles are stored on the local disk, the process is a bit faster but still took too long to be considered for a clinical setting. We recommend that the model be modified to make virtual tiles on the WSI and analyse straight. Annotation quality – the model does not draw annotations or indicate the specific area the AFB is located on the image (tile) making the review by the Pathologist much difficult and time consuming. Furthermore, the authors did not take into consideration artifacts that can be mistaken as AFBs.

Table 2. Total number of tiles generated from Scripts #2 and #3 with their corresponding number of Positive tiles from Model A analysis

SID	Tiles from *.ndpi				Tiles from *.svs			
	Script #2		Script #3		Script #2		Script #3	
	Total	Positive	Total	Positive	Total	Positive	Total	Positive
S01	10,710.00	792.00	169,932.00	37,775.00	6,348.00	2,357.00	101,568.00	20,389.00
S02	9,954.00	442.00	158,760.00	35,920.00	5,856.00	3,766.00	92,062.00	17,194.00
S03	9,717.00	290.00	154,350.00	30,618.00	5,978.00	3,871.00	94,380.00	24,164.00
S04	8,800.00	228.00	139,650.00	27,256.00	5,538.00	3,159.00	87,391.00	15,598.00
S05	11,760.00	2,714.00	188,160.00	36,764.00	6,832.00	2,149.00	107,727.00	44,928.00
S06	4,662.00	2,186.00	74,088.00	31,254.00	2,166.00	553.00	34,200.00	10,019.00
S07	10,440.00	1,021.00	164,934.00	16,617.00	6,840.00	809.00	108,720.00	23,046.00
S08	9,180.00	1,889.00	144,942.00	29,766.00	5,494.00	1,476.00	86,775.00	23,191.00
S09	6,399.00	1,812.00	101,430.00	27,481.00	3,000.00	264.00	48,000.00	19,544.00
S10	10,355.00	1,384.00	164,052.00	21,488.00	6,264.00	939.00	98,610.00	29,990.00
S11	9,310.00	3,310.00	148,176.00	27,181.00	7,144.00	2,758.00	113,250.00	20,842.00
S12	7,134.00	2,114.00	113,190.00	19,488.00	3,393.00	476.00	53,785.00	10,153.00
S13	20,202.00	8,176.00	321,048.00	66,599.00	5,400.00	1,442.00	85,752.00	18,221.00
S14	12,635.00	2,537.00	201,096.00	32,290.00	8,284.00	1,613.00	130,766.00	33,031.00
S15	8,190.00	1,166.00	129,948.00	7,200.00	5,092.00	832.00	80,901.00	16,567.00
S16	9,975.00	1,334.00	158,760.00	18,048.00	6,800.00	946.00	108,141.00	14,607.00
S17	10,360.00	2,851.00	164,640.00	33,037.00	5,664.00	1,334.00	89,006.00	15,567.00
S18	13,209.00	2,019.00	209,916.00	33,586.00	8,544.00	2,513.00	134,846.00	37,844.00
S19	10,640.00	1,434.00	169,344.00	13,711.00	6,786.00	3,619.00	107,260.00	28,215.00
S20	9,975.00	1,302.00	158,760.00	17,295.00	6,786.00	2,651.00	107,260.00	20,407.00
S21	5,733.00	1,483.00	91,728.00	31,985.00	2,760.00	1,255.00	43,498.00	16,089.00
S22	9,240.00	1,861.00	147,000.00	26,348.00	5,986.00	2,344.00	94,792.00	32,243.00
S23	8,820.00	3,641.00	141,120.00	41,482.00	4,720.00	920.00	74,094.00	17,007.00
S24					6,486.00	2,042.00	102,202.00	13,916.00
Totals	227,400.00	45,986.00	3,615,024.00	663,189.00	138,161.00	44,088.00	2,184,986.00	522,772.00

5.2 Model A

Description. This study was published in 2022 and was conducted in Japan. Although the main author had left the group, the PI and his team were kind enough to agree to a collaboration and shared various versions of the HALO AI classifier that were trained to detect bacilli on ZN stain human tissue. After testing all five versions of the trained classifier, we focused on the one that was used for publication (v3.7.5) and the latest version (v4.3).

Using the same WSI dataset that was produced on two different slide scanner, we tested the model on both imaging formats – *.ndpi and *.svs. First, the image resolution of each WSI had to be set to 1.3 microns per pixel in HALO. The parameters of the classifier v3.7.5 on HALO are Resolution ($\mu\text{m}/\text{px}$) = 0.25, and the Probability Threshold (%) = 90.00. The model was trained using four colour coded distinct classes – AFB (red), Background (white), Artifact1 (yellow), and Artifact2 (blue); while classifier v4.3 has the value of 0.84 for Resolution ($\mu\text{m}/\text{px}$) with the Probability Threshold (%) value set to 80.00 and has only two training classes – AFB (red), Background (white). Both classifiers have Minimum Objects Size (μm^2) parameter set to 0.00 and are set to highlight (heatmap) and annotate the detected AFBs (Fig. 3).

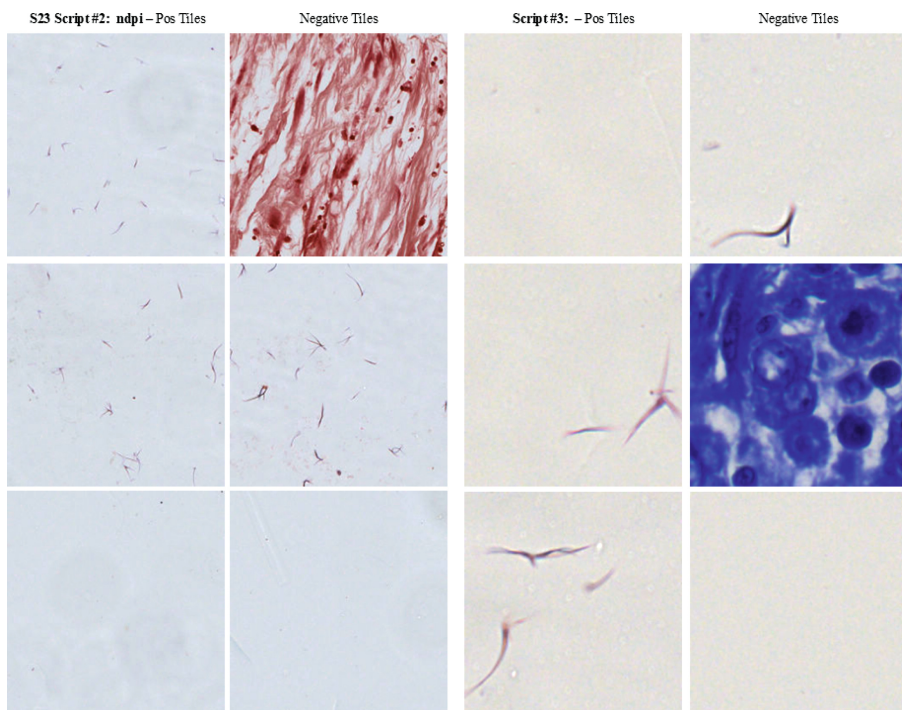


Fig. 3. Model A – sample of prediction S23 ± tiles from two image formats and scripts.

Performance and Limitations. The shared classifiers/models were imported into HALO v3.6.4134.396 and HALO AI 3.6.4134 as compressed folders containing multiple files such as the Classifier file, Params file, and the Trainer file. An annotation was drawn around the stained tissue on the WSI to limit the detection or analysis within the area of interest. After which the classifier was loaded into the platform and the analysis was conducted on Annotation Layer only for each WSI. On completion, models produced an output image overlayed with both – annotations around and a heatmap over the detected AFBs. This allowed us to investigate the correctness of the detection. It also produces quantitative results that can be exported as *.csv file along with the annotated image. Both versions (v3.7.5 and v4.3) annotates artifacts as AFBs. The process seems to be taking longer on *.svs compared to the *.ndpi. The v4.3 classifier performs slightly faster than the v3.7.5. When the option to draw annotations on detected AFB is set to true, the classifiers take even longer to complete regardless of which image format is used (Figs. 4 and 5).

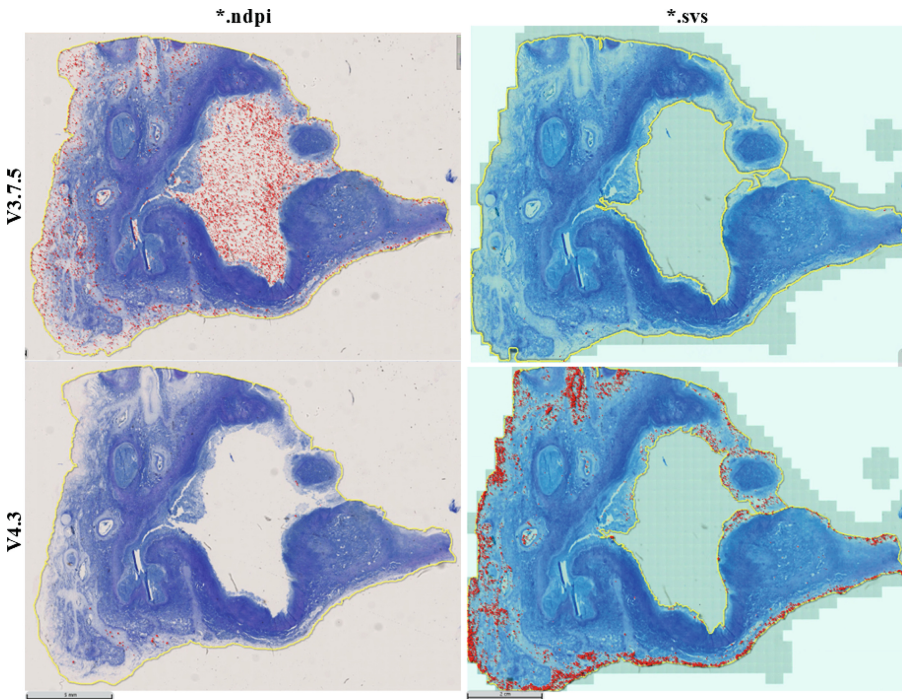


Fig. 4. Model B prediction on S01 from two whole slide imaging formats.

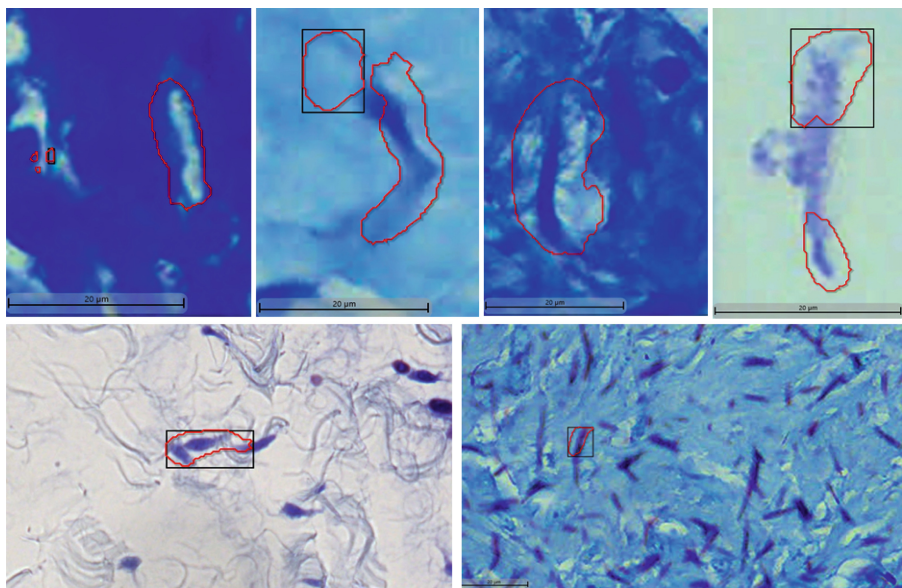


Fig. 5. Model B – examples of detected AFBs from different WSIs.

6 Comparative Analysis

For the comparison of the machine learning models, we focus on how each model can integrate into practical workflow and how easily each model's result can be interpreted. Evaluating how well each model can handle variations in stain quality, sample types, and image conditions. Furthermore, we look into the computational efficiency of each model as this is a crucial factor that impacts practical applications.

6.1 Practical Application

Taking into consideration factors such as ease of deployment in clinical settings, required computational resources, and compatibility with existing laboratory equipment, we observed both models on our internal settings using the virtual machine platform, file server and local PC.

Model A required a huge storage space as it produced more output images (tiles). The total size of tiles for each WSI is bigger than the corresponding WSI for both image modalities when Script #3 is used (Table 3). From our small dataset, the average WSI size in kilobytes (KB) for the *.svs images is 897,902.67 with tiles average size of 421,401.60 from script #2 and 5,611,192.32 from script #3; while the same for the *.ndpi files are 2,063,667.74 with 874,015.17 average kilobyte from script #2 as depicted on Table 3. This alone makes it almost impossible for clinical settings since most of them in the developing world can't afford large computational storage to keep up with the large output. It took hours to process each WSI through the model pipeline – read, split, set parameters, analyse, separate, review on either of the platforms; however, it took even

longer when accessing WSIs and saving the split tiles remotely (from the file server on the same network) than using the local storage drive. This affects the time required to diagnose a single case.

Model B has advantages of ease deployment as it depends on an already built platform (HALO AI) which however requires an advance and very expensive computational resources. Using the strength of HALO platform, we are able to access the WSIs either remotely or locally with ease without affecting the processing time. Minimal outputs are generated which can be reviewed within the platform overlaying the input image. Furthermore, using the power of HALO AI, the model can continue to learn from new WSI making it more efficient and accurate.

6.2 Interpretation

Here we assess how each model's predictions can be understood and verified by clinicians. Does the model provide visual explanations or confidence scores that aid in decision making? We further discuss how transparent the model's decision-making process is as this impacts trust and usability. Also, we delve into how each model supports real-time feedback and error correction, which is crucial for accurate diagnostics.

To improve the work of the Pathologist in a clinical setting, the whole pipeline must be faster from the digitalisation of the slides to the prediction results. Furthermore, they must provide clear distinction between area with AFBs and those without. **Model A**, as received, consumes more time through the pipeline and does not indicate where on the image (WSI or tiles) the AFB is found either by way of annotating or highlighting the bacilli. **Model B** on the other hand highlights the detected AFBs (whether true or false positive) with an option to draw an annotation around them, which adds on processing time. Using the features within HALO, we can switch on and off the overlay to examine and distinguish the true positive from the false. As one can imagine, this exercise can be time consuming when there are thousands of detections on single WSI of which the majority are false positive (Fig. 6).

6.3 Interpretation

When comparing machine learning models, computational efficiency is a critical factor that impacts practical applications and it encompasses processing time, resource utilization, scalability, and power consumption. Balancing these factors ensures that the trained model not only performs well but is also practical and cost-effective for real-world applications.

Processing Time – faster models can handle higher volumes of samples and the model with faster inference is more suitable for real-time applications. As indicated earlier, both models take too long to process a WSI for AFB predictions, though Model B is slightly better since it is based on an already established platform (HALO, Indica Lab). Improving processing time might fast track the use of machine learning models in clinical settings for tuberculosis.

Resource Utilization – models that require less computational power can be deployed on less expensive or existing hardware, reducing costs. Conversely, models that are resource-intensive might demand high-performance servers or cloud computing

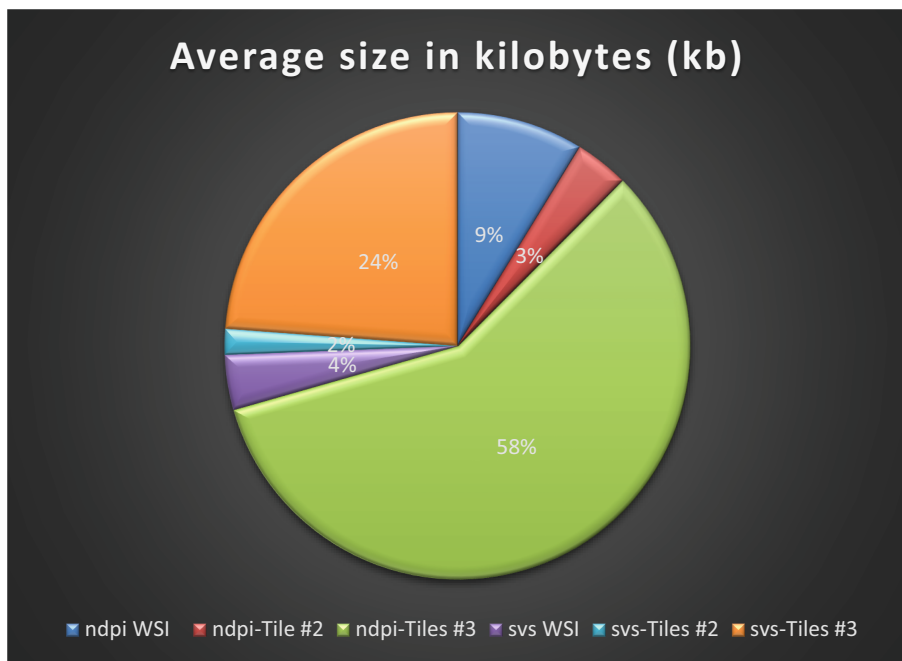


Fig. 6. Average size in Kilobytes (KB) of each WSI format and generated tiles per Script.

resources, which could be a limiting factor in resource-constrained environments. Beside needing huge storage space, Model A requires less computational power than Model B, but it needs to be redesigned in such a way the need to use large storage capacity is cut. Model B runs on licensed software that requires enormous computational power. Its AI functionality does not run if the system on which it is installed does not have a graphic card.

Scalability – a model that performs efficiently with small datasets but struggles as data volume grows might not be suitable for large-scale deployment. Scalability ensures that as the volume of samples increases, the model can handle the load without significant degradation in performance or requiring disproportionate resources. The same can also be said that a model that struggles with small datasets will probably struggle with large volumes. Unfortunately, both models did not perform as expected with a very small dataset of 24 WSI.

Power Consumption – models that are more computationally efficient often consume less energy, which is beneficial for mobile devices or low-power systems. In clinical settings, specialists are often on the move and depend on portable devices to access or provide diagnosis results or treatment guidelines. It is important that ML models be able to run on such devices. At this stage though, none of the models A and B are compatible for portable devices. However, a portable device can be used to access the virtual machine and run the models as needed. In our case we used a Samsung Galaxy Tab S8 to access the virtual machine via VMware Horizon Client (version 2406, build 8.13.0).

Table 3. Size in Kilobytes (KB) of each WSI format with total size of generated tiles per Script

	*.ndpi			*.svs		
	WSI	Script #2	Script #3	WSI	Scrip #2	Script #3
S01	2,489,457.00	1,027,604.48	15,728,640.00	1,108,964.00	445,440.00	7,035,944.96
S02	2,073,487.00	857,088.00	14,155,776.00	814,186.00	372,736.00	5,567,938.56
S03	1,960,655.00	817,152.00	13,526,630.40	773,751.00	368,640.00	5,536,481.28
S04	2,030,583.00	860,160.00	13,002,342.40	808,018.00	375,808.00	5,599,395.84
S05	2,584,207.00	1,111,490.56	16,672,358.40	1,529,413.00	632,832.00	9,709,813.76
S06	887,848.00	340,992.00	6,354,370.56	344,899.00	651,264.00	2,128,609.28
S07	2,677,517.00	1,121,976.32	15,623,782.40	1,591,296.00	131,072.00	9,846,128.64
S08	2,058,773.00	863,232.00	13,107,200.00	1,018,556.00	441,344.00	6,574,571.52
S09	1,419,838.00	566,272.00	9,227,468.80	178,390.00	54,886.40	1,038,090.24
S10	2,623,998.00	1,121,976.32	15,518,924.80	1,066,595.00	552,960.00	7,371,489.28
S11	1,849,068.00	785,408.00	12,687,769.60	881,640.00	382,976.00	5,651,824.64
S12	1,500,669.00	610,304.00	9,930,014.72	563,480.00	270,336.00	3,753,902.08
S13	3,302,536.00	1,216,348.16	25,690,112.00	760,598.00	314,368.00	4,687,134.72
S14	2,569,114.00	1,111,490.56	17,091,788.80	1,129,274.00	553,984.00	7,633,633.28
S15	1,585,354.00	753,664.00	10,359,930.88	756,660.00	428,032.00	5,274,337.28
S16	2,017,607.00	945,152.00	13,002,342.40	896,215.00	529,408.00	6,322,913.28
S17	2,242,721.00	934,912.00	14,575,206.40	1,058,763.00	464,896.00	6,815,744.00
S18	2,825,747.00	1,205,862.40	18,350,080.00	1,168,718.00	630,784.00	8,430,551.04
S19	2,204,917.00	992,256.00	14,575,206.40	910,550.00	517,120.00	6,941,573.12
S20	2,170,255.00	1,006,592.00	13,946,060.80	979,784.00	547,840.00	7,361,003.52
S21	933,698.00	373,760.00	7,423,918.08	212,681.00	103,424.00	1,520,435.20
S22	1,651,295.00	702,464.00	12,163,481.60	584,645.00	295,936.00	4,173,332.48
S23	1,805,014.00	776,192.00	12,058,624.00	935,206.00	359,424.00	5,693,767.68
S24				1,477,382.00	688,128.00	–
Ave	2,063,667.74	874,015.17	13,685,740.41	897,902.67	421,401.60	5,611,192.32
Total	47,464,358.00	20,102,348.80	314,772,029.44	21,549,664.00	10,113,638.40	134,668,615.68

Deployment Considerations – models that require less computational power are easier to integrate into existing systems without extensive modifications.

7 Results and Discussion

7.1 Evaluation of Models

To evaluate the performance of the ML model, the following metrics are often used:

- Accuracy – the proportion of correctly classified AFB versus non-AFB instances.

- Sensitivity (Recall) – the ability of the model to correctly identify AFB cases. High sensitivity is crucial for detecting true positives.
- Specificity – the ability of the model to correctly identify non-AFB cases. High specificity helps in reducing false positives.
- Precision – the proportion of true positive AFB detections out of all detections made by the model.
- F1 Score – the harmonic mean of precision and recall, providing a balance between the two metrics.
- Area under the receiver operating characteristic curve (AUC-ROC) – indicated the model's ability to distinguish AFBs and non-AFBs.
- Area under the precision-recall curve (AUC-PR) – is useful when dealing with imbalanced datasets.

Considering that the above-mentioned metrics were used on both models before their publications, we did consider not recalculate these metrics. Our aim was to test the usability of the model and if they can perform each based on the published evaluation of metric values. However, due to the time constraints it would not be feasible to go through the thousands of tiles generated from Model A just to differentiate the true from false positive tiles. Though Model B generated fewer annotations relative to Model A tiles number; majority, if not all the annotations appear to be false positive. Table 4 contains the total number of annotations per WSI generated by Model B. The annotated images and analysis data (*.csv) are available GitHub (<https://github.com/lumamkd/Comparative-Analysis-of-AI-based-Algorithms>).

Table 4. Total number of annotations by WSI generated from Model B using v4.3 and v3.7.5

	*.ndpi		*.svs	
	v4.3	v3.7.5	v4.3	v3.7.5
S01	62	62	33327	32
S02	66	66	267	73
S03	217	217	280	63
S04	153	153	204	9
S05	102	102	53	52
S06	0	0	3463	15
S07	44	44	35	16
S08	0	0	38	13
S09	1	1	-	-
S10	0	0	24	51
S11	552	552	72	554
S12	0	0	0	55

(continued)

Table 4. (continued)

	*.ndpi		*.svs	
	v4.3	v3.7.5	v4.3	v3.7.5
S13	104	104	24	243
S14	3	3	10	66
S15	212	212	2	41
S16	14	14	10	82
S17	216	216	11	111
S18	14	14	647	140
S19	15	15	15836	205
S20	3	3	585	196
S21	1	1	1123	33
S22	8	8	4138	67
S23	2	2	281	0
S24	–	–		24
Total	1789	1789	60430	2141

7.2 Evaluation of Models

(See Table 5)

7.3 Evaluation of Models

Conducting a comparative study of two machine learning models for detecting Acid-Fast Bacilli on Whole Slide Images has significant implications for both the fields of tuberculosis diagnostics and computer science.

Implications for Tuberculosis Diagnosis

- Improved Diagnosis Accuracy – effective ML models can enhance the accuracy of AFB detection, leading to better TB diagnosis and management. Accurate detection is crucial for timely treatment and containment of TB, which is a global health challenge.
- Reduced Diagnosis Time – automation through ML can potentially reduce the time required for diagnosing TB from WSI, improving efficiency in clinical settings and reducing the burden on pathologists.
- Standardisation – ML models can provide a standardised approach to AFB detection, reducing variability and subjectivity associated with manual examination by pathologists.

Table 5. Models' results interpretations

	Model A	Model B
Performance	Positive Predictions: Model A identified many tiles as positive for AFB. However, a significant portion of these tiles did not contain AFBs upon closer examination. This indicates that while the model is good at identifying potential regions of interest, it suffers from a high false-positive rate Tile Production and Storage: The model's process of splitting WSIs into tiles and analyzing them requires considerable storage and memory, which can be problematic. The large number of tiles produced contributes to the high storage demands	Annotations and Highlights: Model B, using the HALO AI platform, provides visual annotations and heatmaps around detected AFBs. This feature improves interpretability, allowing pathologists to easily review and verify detected AFBs. However, the model also annotates artifacts as AFBs, indicating that it sometimes mistakes non-bacillary features for AFBs
Computational Efficiency	Processing Time: The model's pipeline is quite time-consuming, often taking over hours per WSI. This is primarily due to the high resolution and tile size requirements, which strain computational resources. This extended processing time makes it impractical for real-time clinical use	Processing Time: Although Model B is integrated into the HALO platform, which streamlines some processes, it still has lengthy processing times, particularly when annotation features are enabled. This suggests that while the model is somewhat more efficient than Model A, it still faces challenges with processing speed
Annotation Quality	Lack of Annotations: One significant drawback is the model's inability to annotate or specify the location of AFBs within the tiles. This lack of visual feedback means pathologists have to manually review large numbers of tiles, which can be inefficient and increase the chances of oversight	Visual Explanations: The ability of Model B to provide annotations and heatmaps enhances its usability in clinical settings. Pathologists can use these visual aids to quickly locate and assess areas with potential AFBs. However, the presence of false positives means that additional verification is necessary, which can be time-consuming

(continued)

Table 5. (continued)

	Model A	Model B
Transferability Constraints	Resource and Practicality Issues: The model's large storage requirements and lengthy processing times limit its applicability in resource-constrained settings. These constraints hinder its practicality in clinical environments, particularly in low-resource settings	
Resource Utilization		Computational Resources: Model B requires substantial computational power due to its dependence on the HALO platform. While this platform allows for more sophisticated analysis and annotations, it also means that the model's deployment might be limited to institutions with high-performance computing resources

Implications for Computer Science

- Advancements in ML Techniques – comparing different models can contribute to the development of more effective algorithms for medical image analysis. Insights gained can lead to improvements in design and training of models.
- Integration of AI in healthcare – there is potential of integrating AI into healthcare workflows considering its effectiveness and reliability.

Model generalization – our findings can offer insight into how well different models generalise across diverse datasets and conditions, which is crucial for developing robust ML solutions.

8 Summary of Key Findings

8.1 Evaluation of Models

Dataset Expansion. Diversification of datasets – expand the dataset to include a wide variety of cases from different populations, geographic regions, and stages of TB. This helps in improving the generalisability of the models. We are currently building such a dataset and would like to appeal to different groups around the world to reach out and contribute with WSI to advance the development of machine learning in the TB field. For

Models A and B, increase dataset diversity and size to better evaluate their performance across different conditions and settings. Also implement more comprehensive evaluation metrics that consider the practical aspects of model deployment in clinical settings (Table 6).

Table 6. Summary of key findings

	Model A	Model B
Description and Setup	Employs a CNN-based active learning framework for AFB detection in ZN-stained tissue samples The model was trained on tiles extracted from WSIs, with specific resolution and tile size requirements Issues arose during the tile extraction process due to high memory consumption when setting the resolution as recommended	Uses the HALO AI platform to detect AFBs in ZN-stained tissue samples Configured to operate at a higher resolution and used a platform-based approach for analysis and annotation
Performance and Limitations	The model identified many tiles as positive for AFB, but upon closer examination, many of these tiles did not contain actual AFBs Processing a single WSI through the full pipeline was time-consuming, which is impractical for clinical settings The model does not provide specific annotations for AFB locations The model generated a large number of output tiles, consuming storage space	Effectively produced annotations and heatmaps for detected AFBs, although it also identified artifacts as AFBs Showed better performance compared to model A, though processing times remained lengthy, particularly with annotation features enabled Provided annotations around detected AFBs, improving interpretability for the pathologists, though false positive was also annotated
Practical Application	Large storage requirements and slow processing times make it less practical for clinical settings. The model's inability to annotate AFB locations adds to the review complexity	Offers better deployment options due to integration with the HALO platform. The platform's ability to annotate and highlight AFBs aids in interpretation but still requires substantial computational resources

(continued)

Table 6. (continued)

	Model A	Model B
Interpretability	Lack of visual annotations makes interpretation challenging. The model's results are less transparent, requiring additional effort for validation	Provides visual annotations and heatmaps, enhancing interpretability and facilitating pathologist review, despite some false positives
Computational Efficiency	Consumes less computational power but demands large storage capacity, which is a significant drawback for clinical use	Runs on a powerful platform with high computational requirements but benefits from established software and supports continued learning

Model Improvement. Algorithmic innovations – investigate newer ML algorithms or architectures, such as deep learning models or ensemble methods, to enhance performance further. Feature engineering – explore additional features or preprocessing techniques that could improve the models' ability to detect AFB efficiently. For Model A, optimize memory usage and reduce storage requirements by improving the tile extraction process or using virtual tiles (mask). For Model B, continue refining the model to reduce false positives and improve processing efficiency.

Clinical Integration. Pilot studies – conduct pilot studies to evaluate how well the models integrate into existing diagnostic workflows and their impact on clinical decision making. User feedback – gather feedback from pathologists and clinicians on the models' usability and performance in a real-world environment. For models A and B, develop strategies to enhance the real-time processing capability to make them more suitable for clinical environments. Further, explore ways to improve model portability and integration with existing clinical workflows.

Ethical and Practical Considerations. Bias and Fairness – it is important to ensure that the models do not introduce or perpetuate biases, and we must take into consideration how they might affect different demographic groups. *Regulatory compliance* – ensure there that the models meet regulatory requirements and set guidelines for medical diagnostics tools.

References

1. Ahuja, R.K., Kumar, M.T., Singh, R.N., et al.: Challenges in using small sample sizes for training ai algorithms: a case study in tuberculosis diagnosis. *IEEE Trans. Biomed. Eng.* **67**(1), 123–131 (2020)
2. Anderson, T.J., Patel, H.M., Stewart, A.C., et al.: the gap between annotation datasets and real-world data in tuberculosis diagnostics: challenges and opportunities. *J. Med. Syst.* **45**(44) (2021). <https://doi.org/10.1007/s10916-020-01720-1>

3. Azarafza, M., Ghazifard, A., Akgün, H., et al.: Development of a 2d and 3d computational algorithm for discontinuity structural geometry identification by artificial intelligence based on image processing techniques. *Bull. Eng. Geol. Environ.* **78**(5), 3371–3383 (2019). <https://doi.org/10.1007/s10064-018-1298-2>
4. Babatunde, O.H., Armstrong, L., Leng, J., et al.: A survey of computer-based vision systems for automatic identification of plant species. *J. Agric. Inform.* **6**(1) (2015). <https://doi.org/10.17700/jai.2015.6.1.152>
5. Barberis, I., Bragazzi, N.L., Galluzzo, L., et al.: The history of tuberculosis: from the first historical records to the isolation of koch's bacillus. *J. Prev. Med. Hyg.* **58**(1), E9–E12 (2017)
6. Bharati, S., Podder, P., Rubaiyat, M., et al.: Medical imaging with deep learning for covid-19 diagnosis: a comprehensive review (2021)
7. Chen, Y., Janowczyk, A., Madabhushi, A.: Quantitative assessment of the effects of compression on deep learning in digital pathology image analysis. *JCO Clin. Cancer Inform.* **4**, 221–233 (2020). <https://doi.org/10.1200/CCI.19>
8. Chen, P.R., Liu, Y.Z., Xu, X.W., et al.: Expanding tuberculosis dataset scope: addressing the limitations of focusing on a single disease subtype. *BMC Med. Inform. Dec. Mak.* **21**(30) (2021). <https://doi.org/10.1186/s12911-021-01440-4>
9. Chithra, P.L., Bhavani, P.: A study on various image processing techniques. *Int. J. Emerg. Technol. Innov. Eng.* **5**(5) (2019)
10. Daniel, T.M.: The history of tuberculosis. *Respir. Med.* **100**(11), 1862–1870 (2006). <https://doi.org/10.1016/j.rmed.2006.08.006>
11. Hinton, G.E., Srivastava, N., Krizhevsky, A., et al.: Improving neural networks by preventing co-adaptation of feature detectors (2012). <http://arxiv.org/abs/1207.0580>. Accessed on 11 Nov 2023
12. Hunter, R.L.: Tuberculosis as a three-act play: a new paradigm for the pathogenesis of pulmonary tuberculosis. *Tuberculosis* **97**(8–17) (2016). <https://doi.org/10.1016/j.tube.2015.11.010>
13. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks (2012). <http://code.google.com/p/cuda-convnet>. Accessed on 11 Nov 2023
14. Krogh, A.: What are artificial neural networks? *Nat. Biotechnol.* **26**(2) (2008). <https://doi.org/10.1038/nbt1386>
15. Lee, S.H., Mo, A.N., Kim, Y.W., et al.: Addressing bias and diversity issues in AI-based medical imaging: lessons from tuberculosis datasets. *J. Artif. Intell. Res.* **67**, 455–470 (2020). <https://doi.org/10.1613/jair.1.11754>
16. Leung, A.N.: Pulmonary tuberculosis: the essentials. *Radiology* **210**(2), 307–322 (1999)
17. Lopez, L.J., Rahman, S.T., Hassan, M.M., et al.: Whole slide image dataset of tuberculosis and non-tuberculosis lung samples with cellular annotations. *Sci. Reports* **11**(2651) (2021). <https://doi.org/10.1038/s41598-021-82382-0>
18. Lumamba, K.D., Wells, G., Naicker, D., Naidoo, T., Steyn, A.J.C., Gwetu, M.: Computer vision applications for the detection or analysis of tuberculosis using digitised human lung tissue images – a systematic review. *BMC Med. Imaging* (2024). <https://doi.org/10.1186/s12880-024-01443-w>
19. Madabhushi, A., Lee, G.: Image analysis and machine learning in digital pathology: c and opportunities. *Med. Image Anal.* **33**, 170–175 (2016). <https://doi.org/10.1016/j.media.2016.06.037>
20. Myles, A.J., Feudale, R.N., Liu, Y., et al.: An introduction to decision tree modeling. *J. Chemom.* **18**(6), 275–285 (2004). <https://doi.org/10.1002/cem.873>
21. Noble, W.S.: What is a support vector machine? *Nat. Biotechnol.* **24**(12), 1565–1567 (2006). <https://doi.org/10.1038/nbt1206-1565>

22. Perkins, M.L., Hill, C.A., Magee, J.F., et al.: Annotated whole slide image dataset for tuberculosis diagnosis using deep learning. *J. Pathol. Inform.* **11**(1), 10–22 (2020). https://doi.org/10.4103/jpi.jpi_37_19
23. Razzak, M.I., Naz, S., Zaib, A.: Deep learning for medical image processing: overview, challenges and the future. *Lect. Notes Comput. Vis. Biomech.* **26**, 323–350 (2018). <https://doi.org/10.1007/978-3-319-65981-712>
24. Smith, V.M., Thompson, L.J., Johnson, B.K., et al.: Standardizing tuberculosis whole slide imaging: current practice and future directions. *Dig. Pathol.* **19**, 55–62 (2020). <https://doi.org/10.1016/j.dpath.2019.12.004>
25. Suthaharan, S.: Support vector machine. In: *Integrated Series in Information Systems*, vol. 36, p 207–235. Springer, Boston, MA (2016). <https://doi.org/10.1007/978-1-4899-7641-3>
26. Tajbakhsh, N., Jeyaseelan, L., Li, Q., et al.: Embracing imperfect datasets: a review of deep learning solutions for medical image segmentation. *Med. Image Anal.* **63** (2020). <https://doi.org/10.1016/j.media.2020.101693>
27. Umaa Mageswari, S., Sri devi, M., Mala, C.: An experimental study and analysis of different image segmentation techniques. *Procedia Eng.* **64**, 36–45 (2013). <https://doi.org/10.1016/j.proeng.2013.09.074>
28. WHO. Global tuberculosis report 2023 (2023). <https://www.who.int/tb/publications/global-report/en>. Accessed on 5 June 2024
29. Xu, F., Wong, M.B., McGill, G.B., et al.: A large-scale whole slide image dataset of tuberculosis and non-tuberculosis lung tissue for the development of AI-based diagnostic tools. *J. Digit. Imaging* **33**(2), 399–407 (2020)
30. Yang, M., Nurzynska, K., Walts, A.E., et al.: A CNN-based active learning framework to identify mycobacteria in digitized ziehl-neelsen stained human tissues. *Comput. Med. Imaging Graph.* **84** (2020). <https://doi.org/10.1016/j.compmedimag.2020.101752>
31. Yasar, A., Saritas, I., Korkmaz, H.: Computer-aided diagnosis system for detection of stomach cancer with image processing techniques. *J. Med. Syst.* **43**(4) (2019). <https://doi.org/10.1007/s10916-019-1203-y>
32. Zaizen, Y., Kanahori, Y., Ishijima, S., et al.: Deep-learning-aided detection of mycobacteria in pathology specimens increases the sensitivity in early diagnosis of pulmonary tuberculosis compared with bacteriology tests. *Diagnostics* **12**(3) (2022). <https://doi.org/10.3390/diagnostics12030709>
33. Zhang, Z.: Artificial neural network. *Multivariate Time Series Analysis in Climate and Environmental Research*, p. 1–35 (2018). https://doi.org/10.1007/978-3-319-67340-0_1
34. Zhang, J., Zhang, X., Zhang, L., et al.: A comprehensive dataset for tuberculosis research with annotated nuclei and mitotic figures. *Comput. Biol. Med.* **115**, 103–110 (2020)



Classical vs Deep Learning Models: A Study on Maize Leaf Diseases Classification

Chimango Nyasulu^{1,2} , Cecilia Tivir¹ , and Jonathan Mwaura³ 

¹ LANI (Laboratoire d'Analyse Numérique et Informatique), University of Gaston Berger, BP:234, Saint-Louis, Senegal
tivir.cecilia-rafael-jose@ugb.edu

² Department of Information and Communication Technology, Mzuzu University, P/Bag 201, Luwingu, Mzuzu 2, Malawi
nyasulu.chimango@ugb.edu.sn

³ Roux Institute at Northeastern University, Portland, ME 04101, USA
j.mwaura@northeastern.edu

Abstract. The early detection and accurate classification of diseases affecting maize crops remains a significant challenge for farmers, particularly in sub-Saharan Africa. Many farmers still rely on traditional methods to identify diseases, which are often inefficient and inaccurate. To address these challenges, this study investigates the use of classical machine learning models - Decision Tree, Logistic Regression, Random Forest, and eXtreme Gradient Boosting as well as pre-trained Convolutional Neural Network (CNN) models, including VGG16, ResNet50, Xception, and DenseNet201, to classify two common maize leaf diseases: Maize Streak Virus and Fall Armyworm, which are prevalent in Sub-Saharan Africa. The findings reveal that among the classical machine learning models, eXtreme Gradient Boosting achieves the highest accuracy at 73%, while DenseNet201 outperforms all models with an impressive accuracy of 98% among the pre-trained CNN models. Overall, the pre-trained CNN models not only deliver superior results but also significantly reduce the time and effort required for training.

Keywords: Classification · Machine Learning · Convolutional Neural Network · Pre-trained Models · Maize Leaf Diseases

1 Introduction

One of the most significant staple foods for millions of people in Africa is maize [1]. Furthermore, maize plays a substantial role in the economies of many African nations, is a major cash crop, and is frequently the focus of agricultural policies meant to promote food security and reduce poverty [2]. However, maize crops in

Supported by the United States Department of State (DoS): Embassy in Dakar, under the University Partnership Initiative, Grant No. SSG20023GR0032.

Africa are vulnerable to a range of pests and diseases that can negatively affect yields and consequently make people more susceptible to food insecurity [3].

The early detection and accurate classification of diseases affecting maize crops has been one of the many challenges faced by farmers, especially in sub-Saharan Africa [4]. Usually, these farmers resort to traditional techniques to identify diseases. These traditional techniques often rely on visual crop inspections and require expert knowledge, regular and strict monitoring, and generally take a lengthy period to identify and mitigate. Moreover, techniques relying on visual knowledge are often not precise and accurate. This problem becomes more pronounced where large-scale farming is envisioned [5].

Classical Machine Learning (ML) methods serve as a solid foundation for tackling classification tasks. However, when applied to image data, these methods require pre-processing steps to extract feature vectors essential for model training. In contrast, Deep Learning (DL) techniques, such as Convolutional Neural Networks (CNNs), combine feature extraction and learning into a unified framework. This allows DL models to automatically extract intricate features and identify patterns effectively. Both classical ML and DL approaches have been widely employed in the literature for image identification and classification across diverse applications. For instance [6] has used both ML and DL techniques for crop disease detection and classification, while [15] and [16] have used DL for maize leaf disease classifications and [14] as well as [4] have used classical ML for maize leaf diseases classification.

The work reported in this article follows the work reported on [15] and [16] closely. However, while their work utilizes deep learning techniques and reports on their efficiency, no effort is made to investigate classical ML techniques. In addition, the reported work is inspired by work reported in [14] and [4], although their work only concentrates on classical ML techniques with no efforts to compare with DL techniques. Further, whereas the work reported in [6] compares DL and classical ML techniques, they do so on general crop diseases. In comparison, the work reported here is concerned not only with the comparison of classical ML and DL in maize leaf disease classification but also with using a dataset containing maize leaf diseases prevalent in Sub-Saharan Africa. As such, the reported work has an impact on maize leaf diseases identification and classification in Sub-Saharan Africa and also contributes to the broader ML community by offering an empirical comparison between classical ML and DL in maize leaf diseases classification.

The rest of the paper is organized as follows: Sect. 2 presents a brief but detailed overview of related work, while Sect. 3 presents the study's methodology. In Sect. 4, the results and discussion of the study's findings are presented. Finally, Sect. 5 concludes and summarizes the study.

2 Related Work

This section explores previous, current, and ongoing approaches to detecting and classifying maize leaf diseases. It starts by providing a brief but detailed

overview of maize crop diseases. Then, it focuses on the various applications of classical and deep learning ML techniques in maize disease classification.

2.1 Maize Crop Diseases

A wide range of diseases can impact the growth, yield, and quality of maize crops. Diseases can affect maize crops at different stages of development and in various parts of the plant, including the leaves, stalks, roots, cobs, kernels, flowers, and tassels. Table 1 presents some of the prevalent diseases that affect maize in Africa. A closer look in Table 1 suggests that a majority of maize-related diseases affect the leaves of the crop. Consequently, the research reported in this work investigated maize leaf diseases as these are the majority and have very similar symptoms.

As previously mentioned, maize is a staple crop across the Sub-Saharan Africa. As such, various non-ML-related research studies have investigated maize diseases. A good overview of this research can be found here [2]. The majority of research has focused on identification and treatment from an agricultural viewpoint. The work reported here aims to employ machine learning techniques in this task, as ML models have shown tremendous success in other agricultural-related tasks.

2.2 Review of Related Work

Several researchers have used different ML-based techniques, including classical ML models and pre-trained CNNs, to detect and classify maize leaf diseases.

Panigrahi et al., [14], explored different classical models, namely Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbor (KNN), and Naive Bayes (NB). Among the tested algorithms, the RF outperformed and attained the highest level accuracy of 79.23% in testing image data for disease detection, i.e. whether the leaves are diseased or healthy. Their work shows that Classical ML classification models are suitable tools even when image data is concerned. In general, classical ML techniques are helpful in many applications because of their interpretability, low processing overhead, resilience to overfitting, and quick training and inference.

Subramanian et al., [15] fine-tuned state-of-the-art CNN models, including deep VGG-16, VGG-19, Inception-v3, ResNets, and DenseNets for maize leaf disease detection. The Inception-v3 with the Global Average Pooling layer (Inception-v3_GAP) model achieved the highest accuracy of 95.99%, and it was efficient in learning the relevant features from the disease symptoms, also in predicting correct class levels in the unseen data. A similar study by Haque et al., [16] proposed using a pre-trained CNN model to classify maize leaf diseases: Banded Leaf Maydis Leaf Blight (MLB), Sheath Blight (BLSB), Turcicum Leaf Blight (TLB) and health leaves. The authors used three different deep learning architectures on top of Inception-v3 (global average pooling layer with fully connected layer, global average pooling layer, and flatten layer with fully connected layer). The Inception-v3 model with the global average pooling layer achieved an

Table 1. Some of the prevalent diseases that affect maize in Africa

Reference	Disease Name	Cause	Symptoms
[2]	Maize lethal necrosis	maize chlorotic mottle virus, maize lethal necrosis virus	yellowing and mottling of the older leaves, death of tissue, brown streaks
[7]	Maize Dwarf	Maize Dwarf Virus	stunt plant growth, yellowing of leaves
[8]	Northern Leaf Blight	Exserohilum turcicum	elongated lesions on the leaves
[9]	Fusarium Ear Rot	Fusarium verticillioides	maize ears, poor grain quality, mycotoxin production
[2]	Fall Armyworm	Spodoptera frugiperda	window-like feeding damage on leaves, ragged edges on leaves
[10]	Common Rust	Puccinia sorghi	reddish-brown pustules on leaves
[11]	Gray Leaf Spot	Cercospora zeaemaydis	dark, rectangular spots on leaves
[12]	Charcoal Rot	Macrophomina phaseolina	wilting, ear drop
[13]	Maize Streak	Maize Streak Virus	white or yellow streaks along the leaf veins, twisted or distorted leaves, stunted growth

overall higher testing accuracy of 95.71% in the 70:15 training-validation data configuration. The class-wise prediction accuracies of the Inception-v3 model with the global average pooling layer were 99% for Healthy, 98% for BLSB, 95% for TLB, and 91% for MLB. In comparison to this work, the work reported in this article compares both classical ML and DL using a dataset containing maize leaf diseases prevalent in Sub-Saharan Africa.

Kirola et al., [6] presented a comparison between ML and CNN approaches. Different ML and deep learning techniques, including LR, SVM, RF, KNN, and CNN, were explored for predicting plant diseases. The result showed that among the classical techniques, RF achieved the highest accuracy of 97.12% whereas the CNN technique had an accuracy of 98.43%. From their results, the CNN model is excellent for disease prediction in plants and more suited for accurate, large-scale agricultural applications. This work is noteworthy as it compares both classical ML and DL techniques. However, they do so using a dataset containing different crop diseases. The work reported in this article does draw inspiration from this work but utilizes different dataset in order to do so.

An important aspect to consider when dealing with deep learning models is that it is not always practical to implement and train them from scratch. Further, these techniques require massive amounts of data for training. As such, pre-trained CNNs become increasingly useful due to reduced training time, transfer learning, improved performance, less data requirement, and access to state-of-the-art architectures.

3 Methods

This section describes the proposed ML methodology and techniques for classifying and detecting maize crop leaf diseases. The proposed method will follow several steps: image acquisition, pre-processing, model training and evaluation.

3.1 Description of the Dataset

Maize streak virus (MSV) and fall armyworm (FAW) are among the predominant maize diseases in Africa. This work leveraged an open-source dataset from Makerere AI Lab [17] and another dataset available on the Harvard Dataverse platform [18]. The data was collected over three periods: (i) December 2021, (ii) February 2022, and (iii) May and June 2022). It provides essential features for object detection and diseases classification of maize leaf. The maize crop disease image dataset has 9,034 images. The images are labeled with three class labels: a) maize streak virus, (b) fall armyworm, and c) health (See Fig. 1). The available dataset was limited to just these two maize diseases. About 2,611 belong to health maize, 3,009 to MSV, and 3,414 to FAW. All classes were balanced before training the models. The dataset was split into 75%, 15% and 10% for training, validation and testing of the models.

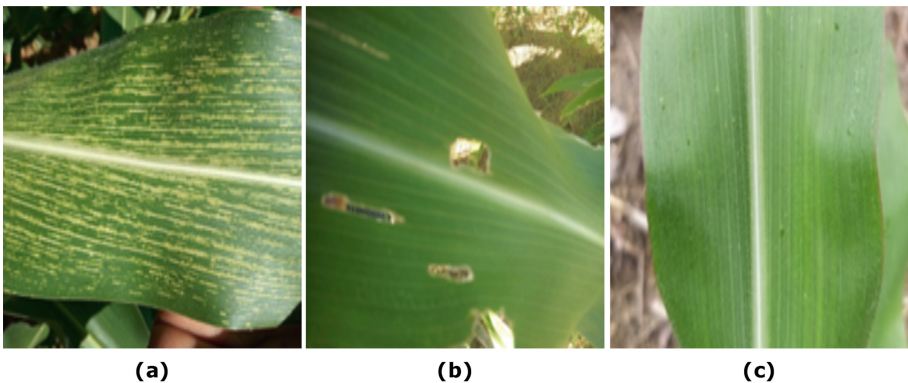


Fig. 1. Sample images of maize leaf disease classes, (a) Maize Streak Virus, (b) Fall Armyworm and (c) Health maize

3.2 Pre-processing of the Dataset

Image pre-processing is crucial when preparing image data for ML tasks, as it greatly enhances the performance of the models. The images were resized to a uniform dimension (224×224) to ensure consistent input shapes across the images. Pixel values were normalized to a range of 0 to 1 to ensure an equal contribution of all input features to the model's classification.

3.3 Machine Learning Models

This study explored the use of classical ML classifiers, including Decision Tree (DT), Random Forest (RF), Logistic Regression (LR) and eXtreme Gradient Boosting (XGBoost). Further, the research used pre-trained CNN models VGG16, ResNet50, Xception, and DensNet201 for the classification of maize leaf diseases. The pre-trained CNN models were fine-tuned by placing a global average pooling layer on top of the base model. Hyperparameters for classical models were optimized using a grid search technique. Grid search offers a clear understanding of how different hyperparameters impact model performance and allows for parallel processing to speed up the tuning process [19]. A summary of the parameters for the classical ML models is presented in Table 2.

Table 2. eXtreme Gradient Boosting, Random Forest, Logistic Regression and Decision Tree parameter setting.

Model	Parameters	Values
XGBoost	learning_rate	0.01
	max_depth	5
	max_leaves	None
	objective	multi:softprob
	subsample	0.5
RF	max_depth	7
	max_features	15
	min_samples_leaf	1
	n_estimators	100
LR	multi_class	ovr
	solver	liblinear
DT	criterion	entropy
	min_samples_leaf	1
	max_depth	5
	max_features	10

3.4 Evaluation Metrics

In ML, evaluation metrics play a critical role in determining a model's performance. Accuracy, Precision, Recall, F-score, confusion matrix and Area Under The Curve (AUC) - Receiver Operating Characteristics (ROC) curve are often used to assess the performance of the classification models [20]. Given the following notations, TP represents True Positives, TN True Negatives, FP False Positives, and FN False Negatives mathematically:

Accuracy (1) measures the ratio of a model's correct predictions to its total number of predictions;

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

Precision (2) refers to the ratio of genuine positives to all positive predictions;

$$Precision = \frac{TP}{TP + FP}, \quad (2)$$

Recall (3) is the ratio of true positives to all positive predictions;

$$Recall = \frac{TP}{TP + FN}, \quad (3)$$

F-score (4) represents the harmonic mean of Recall and Precision.

$$F1 - score = 2 \left(\frac{Precision \times Recall}{Precision + Recall} \right). \quad (4)$$

The confusion matrix is a graphical tool that displays the number of TP, TN, FP, and FN. The AUC-ROC curve illustrates how well the model can distinguish between various classes using the False Positive Rate and the True Positive Rate [20].

4 Results and Discussion

Table 3 shows the performance of the ML and DL techniques.

The results show that XGBoost got an overall higher accuracy of 73% among the classical Machine Learning models, and DensNet201 got an overall higher accuracy of 98% among the pre-trained CNN models and all the models explored in this work. The DenseNet201's layered connectivity is intended to overcome the drawbacks of other CNNs by improving performance through factors like enhanced feature propagation, fewer parameters, feature reusability, and overfitting mitigation. Analyzing the performance of the best model (DensNet201), the disease classes MSV and FAW got 98% for Precision, Recall and F1-score. This demonstrates that the disease symptoms can be distinguished by the model, allowing the model to classify cases correctly. In terms of all classes, the Health class received the highest overall score of 99% for precision, recall, and F1-score.

Table 3. Summary of Accuracy, Precision, Recall, and F1-score performance of XGBoost, RF, Logistic Regression, DT, VGG16, ResNet50, Xception, and DensNet201.

Model	Accuracy	Class	Precision	Recall	F1-score
XGBoost	0.73	Fall Armyworm	0.69	0.66	0.68
		Health	0.84	0.90	0.87
		Maize Streak Virus	0.65	0.64	0.65
RF	0.62	Fall Armyworm	0.60	0.50	0.54
		Health	0.72	0.82	0.77
		Maize Streak Virus	0.51	0.53	0.52
LR	0.66	Fall Armyworm	0.62	0.63	0.63
		Health	0.79	0.84	0.81
		Maize Streak Virus	0.54	0.50	0.52
DT	0.55	Fall Armyworm	0.52	0.47	0.49
		Health	0.67	0.68	0.67
		Maize Streak Virus	0.45	0.49	0.47
VGG16	0.97	Fall Armyworm	1.0	0.90	0.95
		Health	1.0	1.0	1.0
		Maize Streak Virus	0.93	1.0	0.96
ResNet50	0.78	Fall Armyworm	0.62	0.83	0.71
		Health	0.84	0.93	0.88
		Maize Streak Virus	0.93	0.65	0.77
Xception	0.97	Fall Armyworm	0.97	0.99	0.98
		Health	0.98	0.97	0.98
		Maize Streak Virus	0.97	0.96	0.96
DensNet201	0.98	Fall Armyworm	0.98	0.98	0.98
		Health	0.99	0.99	0.99
		Maize Streak Virus	0.98	0.98	0.98

This may be explained by the fact that the Health class is simple to identify from other classes, making it easier for a classifier to identify more positive cases.

The confusion matrices in Fig. 2 for DenseNet201 and Fig. 3 for XGBoost were examined in order to gain a deeper understanding and compare the performance of the best CNN model and classical model.

This analysis assisted in pinpointing the precise situations in which the models performed well or failed. The confusion matrices indicate that the XGBoost struggled to classify samples for MSV and FAW. About 93 samples of MSV were wrongly classified as FAW, and about 83 samples of FAW were improperly classified as MSV. This is attributed to the similarity of symptoms for these diseases and the inability of the classical models to extract deep features from the images. Generally, the DenseNet201 model achieved a higher overall number of correct

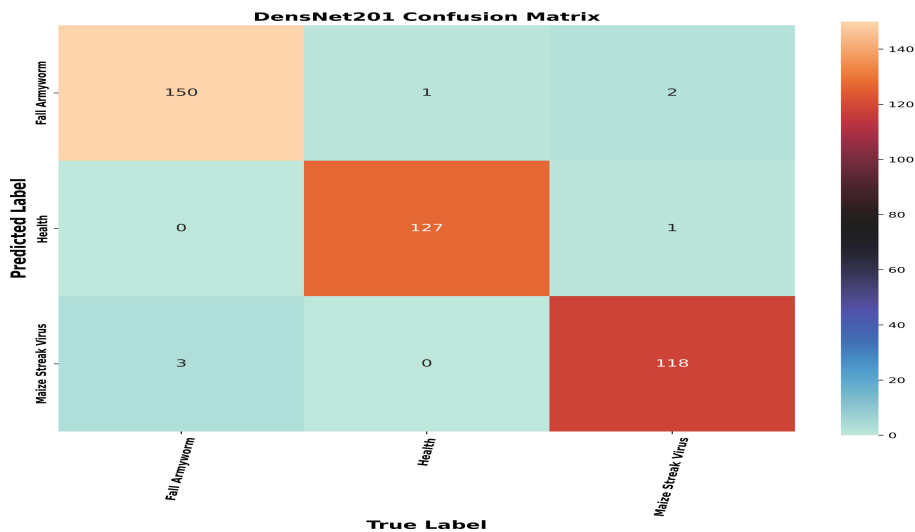


Fig. 2. Confusion matrix of the DenseNet201 showing correct and incorrect classifications.

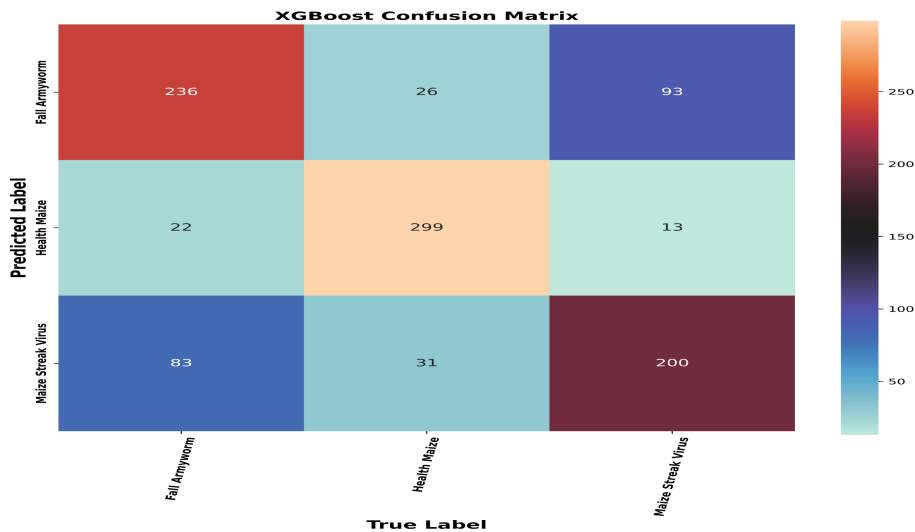


Fig. 3. Confusion matrix of the XGBoost showing correct and incorrect classifications.

classifications because of the ability of the CNN models to extract deep features from the images. Additionally, the AUC - ROC curve was used to visualize how well the DenseNet201 model in Fig. 4 and XGBoost model in Fig. 5 performed. The AUC values for the DenseNet201 model in Fig. 4 show that the model achieved an AUC of 100% for all classes. This indicates the ability of the model to distinguish the diseases. Comparatively, the XGBoost model in Fig. 5 struggled

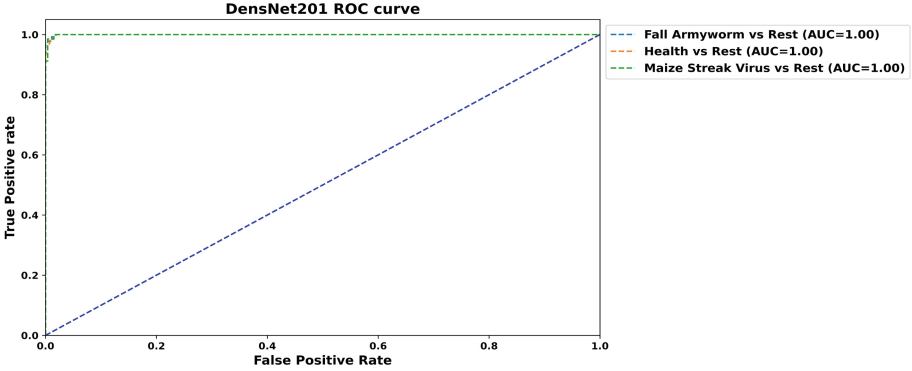


Fig. 4. DenseNet201 Area Under the Curve - Receiver Operating Characteristics curve.

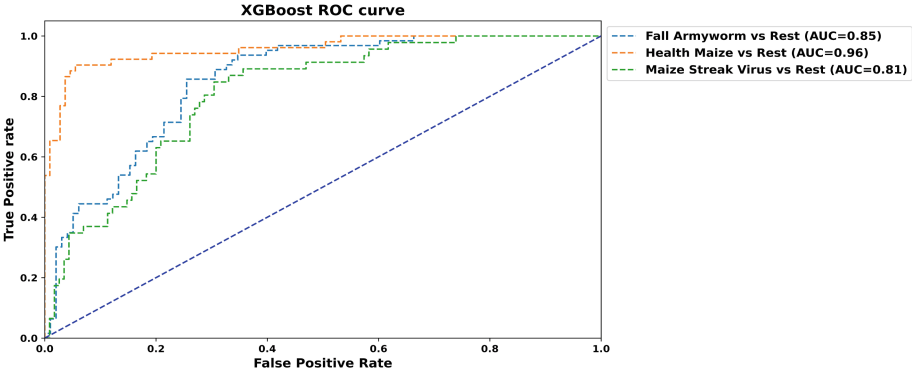


Fig. 5. XGBoost Area Under the Curve - Receiver Operating Characteristics curve.

to distinguish the diseases: 81% AUC for the Maize streak virus class and 85% AUC for the fall armyworm class.

5 Conclusion

The majority of people in sub-Saharan Africa regularly consume maize or maize by-products as their staple food. However, maize is among the crops that are highly susceptible to diseases. These diseases affect maize stems, leaves, roots and the actual product. Most farmers in sub-Saharan Africa resort to traditional techniques to identify crop diseases. Among other limitations, traditional techniques are not accurate and efficient. To contribute to food security, this work explored the use of classical ML models and pre-trained CNN models to classify maize leaf diseases.

Among the models evaluated, DenseNet201 achieved the highest classification accuracy of 98%, demonstrating superior performance over both classical ML methods and other CNN architectures. Generally, the use of pre-trained

CNN models, such as DenseNet201, offers better results in detecting maize leaf diseases. Pre-trained CNNs reduce effort in training models by recognizing general features from millions of images and using smaller datasets, shorter training times, and lower learning rates. In agricultural applications where datasets may be scarce, pre-trained CNNs are especially useful because of their capacity to generalize across a variety of image attributes.

Future work will focus on compiling a comprehensive, region-specific dataset detailing the range of maize leaf diseases prevalent in sub-Saharan Africa. Additionally, efforts will be concentrated on enhancing automated maize leaf diseases detection systems through the use of state-of-the-art ML frameworks. These advancements aim to support farmers in correctly and quickly identifying and treating infections by increasing the precision and reliability of agricultural disease control technology. In the end, it seeks to contribute to improving agricultural resilience and food security across the region by addressing pressing needs for sustainable crop management and rural livelihoods.

References

1. Grote, U., Fasse, A., Nguyen, T.T., Erenstein, O.: Food security and the dynamics of wheat and maize value chains in Africa and Asia. *Front. Sustain. Food Syst.* **4**, 617009 (2021)
2. Njeru, F., Wambua, A., Muge, E., Haesaert, G., Gettemans, J., Misinzo, G.: Major biotic stresses affecting maize production in Kenya and their implications for food security. *PeerJ* **11**, e15685 (2023). <https://doi.org/10.7717/peerj.15685>. PMID: 38050609; PMCID: PMC10693822
3. Onyutha, C.: African food insecurity in a changing climate: the roles of science and policy. *Food Energy Secur.* **8**(1), e00160 (2019)
4. Alehegn, E.: Maize Leaf Diseases Recognition and Classification Based on Imaging and Machine Learning Techniques (Doctoral dissertation) (2020)
5. Flint, M.L.: Pests of the garden and small farm: a grower's guide to using less pesticide, vol. 3332. UCANR Publications (2018)
6. Kirola, M., Joshi, K., Chaudhary, S., Singh, N., Anandaram, H., Gupta, A.: Plants diseases prediction framework: a image-based system using deep learning. In: 2022 IEEE World Conference on Applied Intelligence and Computing (AIC), pp. 307–313. IEEE (2022)
7. Miller, W.A., Lozier, Z.: Yellow dwarf viruses of cereals: taxonomy and molecular mechanisms. *Annu. Rev. Phytopathol.* **60**(1), 121–141 (2022)
8. DeChant, C., et al.: Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. *Phytopathology* **107**(11), 1426–1432 (2017)
9. Lanubile, A., Maschietto, V., Borrelli, V.M., Stagnati, L., Logrieco, A.F., Marocco, A.: Molecular basis of resistance to Fusarium ear rot in maize. *Front. Plant Sci.* **8**, 1774 (2017)
10. Sserumaga, J.P., et al.: Identification and diversity of tropical maize inbred lines with resistance to common rust (*Puccinia sorghi* Schwein). *Crop Sci.* **60**(6), 2971–2989 (2020)
11. Dhami, N.B., Kim, S.K., Paudel, A., Shrestha, J., Rijal, T.R.: A review on threat of gray leaf spot disease of maize in Asia (2015)

12. Darwesh, O.M., Elshahawy, I.E.: Management of sunflower charcoal-rot and maize late-wilt diseases using the aqueous extract of vermicompost (vermitea) and environmental-safe biochar derivative (wood vinegar). *Sci. Rep.* **13**(1), 17387 (2023)
13. Emeraghi, M., Achigan-Dako, E.G., Nwaoguala, C., Oselebe, H.: Maize streak virus research in Africa: an end or a crossroad. *Theor. Appl. Genet.* **134**(12), 3785–3803 (2021). <https://doi.org/10.1007/s00122-021-03914-y>
14. Panigrahi, K.P., Das, H., Sahoo, A.K., Moharana, S.C.: Maize leaf disease detection and classification using machine learning algorithms. In: Das, H., Pattnaik, P.K., Rautaray, S.S., Li, K.-C. (eds.) *Progress in Computing, Analytics and Networking. AISC*, vol. 1119, pp. 659–669. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-2414-1_66
15. Subramanian, M., Shanmugavadeivel, K., Nandhini, P.S.: On fine-tuning deep learning models using transfer learning and hyper-parameters optimization for disease identification in maize leaves. *Neural Comput. Appl.* **34**(16), 13951–13968 (2022)
16. Haque, M.A., et al.: Deep learning-based approach for identification of diseases of maize crop. *Sci. Rep.* **12**(1), 6334 (2022)
17. Namibia University of Science and Technology Maize Dataset Homepage. <https://air.ug/dataset-details/6/>. Accessed 4 Sept 2024
18. Blessing, S., et al.: Namibia University of Science and Technology Maize Dataset (2022). <https://doi.org/10.7910/DVN/6R78HR>. Harvard Dataverse, V2
19. Ndiaye, E., Le, T., Fercoq, O., Salmon, J., Takeuchi, I.: Safe grid search with optimal complexity. In: *International Conference on Machine Learning*, pp. 4771–4780. PMLR (2019)
20. Vujović, Ž: Classification model evaluation metrics. *Int. J. Adv. Comput. Sci. Appl.* **12**(6), 599–606 (2021)



Implementing a Novel Cybersecurity Algorithm for Smart Grid SCADA System

Kholofelo H. Halefose¹, Topside E. Mathonsi¹(✉), Tonderai Muchenje¹,
and Deon Du Plessis²

¹ Tshwane University of Technology, Pretoria, South Africa
{MathonsiTE, MuchenjeT}@tut.ac.za

² University of Johannesburg, Johannesburg, South Africa
danieldep@uj.ac.za

Abstract. The shift to smart grids requires strong cybersecurity measures to protect critical infrastructure from evolving threats. The integration of automation systems increases vulnerability to cyberattacks, particularly targeting Supervisory Control and Data Acquisition (SCADA) systems. Ensuring secure communication and data integrity is essential for maintaining the reliability and resilience of smart grids. This study introduces a novel cybersecurity algorithm for Smart Grid SCADA systems, addressing modern cyber threats. The algorithm enhances authentication, availability, data integrity, and confidentiality while minimizing computational overhead. It integrates N-th degree Truncated polynomial Ring Units (NTRU) and Toom-Cook algorithms for optimized security. The proposed Enhanced Cybersecurity (ECS) algorithm was evaluated using Network Simulator Version-2 (NS-2) against Elliptic Curve Cryptography (ECC) and Rivest-Shamir-Adleman (RSA). Results show high-speed encryption (1.4 ms) and decryption (2.9 ms) with minimal processing resource use. Under DoS attacks, ECS experienced only 0.5% packet loss, proving its suitability for SCADA systems. By securing critical assets, ECS enhances smart grid resilience and dependability, offering a practical solution for strengthening cybersecurity in modern SCADA environments.

Keywords: Smart Grid · Supervisory Control and Data Acquisition (SCADA) · Enhanced Cybersecurity Algorithm (ECS) · NTRU Algorithm · Toom-Cook Algorithm

1 Introduction

Modern Smart Grids with networked smart devices are replacing traditional power grids as we know them. However, this evolution has left the Smart Grid open to security risks and cyberattacks. Given their complexity, Smart Grids are more vulnerable to cyberattacks than the older Traditional Power Grids, whose automation systems were physically isolated from corporate networks. Smart Grid's vulnerabilities to cyber hazards, cyber threats, and cyberattacks are heightened by the Smart Grid's dependence on Supervisory

Control and Data Acquisition (SCADA) systems for automated control and monitoring. Due to these threats, this paper focused on investigating various privacy and SCADA cybersecurity challenges as well as developing strategies, models, and approaches to increase the security and resilience of smart grids [1]. Furthermore, the frequent and ongoing cyberattacks on vital infrastructure show that better cybersecurity solutions are required for smart grids. The current SCADA protocols need to guarantee end-to-end authentication, data integrity, non-repudiation, and confidentiality to fully secure Smart Grid communication [2]. We can see the harm that a determined adversary can do to a nation's security, economy, and healthcare systems thanks to cyberattacks like Stuxnet, Aurora, and Maroochy on SCADA systems [3]. As a result, by safeguarding SCADA systems, we are essentially safeguarding the Smart Grid and, by extension, the nation's vital infrastructure.

The Smart Grid must be secured from cyberattacks since it is regarded as an essential infrastructure for any nation. A successful cyberattack on vital infrastructure carries a high risk because it jeopardises public health and safety, national security, and the nation's economy. A critical security need for NIST and Smart Grids is cybersecurity. Consequently, US Departments in 2012 offered recommendations for implementing cybersecurity in smart grids. These include the three (3) cybersecurity goals of confidentiality, integrity, and availability that need to be met. To strengthen the cybersecurity of the Smart Grid, researchers have tried to identify, create, and suggest cybersecurity algorithms as well as a security plan. On the other hand, new and sophisticated assaults are introduced and released into the internet world at a very fast speed, and cyber attackers are growing more skilled every day. According to CyberEdge's first CyberEdge Report (CDR), which was released in 2014, 62% of organisations had been infiltrated by successful cyberattacks. By 2021, that percentage had increased to 86%. The proportion of organisations that successfully countered an attack increased by 5.5%, marking the biggest growth in the previous six (6) years [4].

It is evident from the frequent and persistent cyberattacks on vital infrastructure that the Smart Grid SCADA system needs better cyber protection. The current SCADA protocols must offer end-to-end authentication, data integrity, confidentiality, and availability in order to completely safeguard the Smart Grid SCADA system communication [5]. Remote access is used to exploit most vulnerabilities in Smart Grid SCADA systems. Furthermore, the SCADA system's weaknesses originate from the current security algorithms' inability to provide real-time monitoring and robust encryption [6]. Moreover, susceptible Smart Grid SCADA systems are targeted by control system viruses like Stuxnet, Aurora, Maroochy, and BlackEnergy. Cybersecurity puts vital infrastructure at risk if it is not implemented effectively. An adversary might take over the entire system and cut off the electricity supply if they were able to breach the Smart Grid SCADA system. People might pass away in hospitals, which would be detrimental to our healthcare systems and the entire nation. The security of our nation will be impacted by this circumstance, which will also bankrupt enterprises and damage our economy financially. As a result, it is critical to defend the Smart Grid SCADA system against hackers because our nation depends on power to function. Thus, the goal of this study is to create an ECS algorithm that will enhance and fortify cybersecurity for the SCADA of the Smart Grid.

1.1 Smart Grid Network Components

Smart Grids rely on two types of communication networks: Home Area Networks (HANs) for local communication within homes and Wide Area Networks (WANs) for broader communication across the grid [7]. The smart meter and other in-home smart devices are connected by a HAN. Bluetooth, cable or wireless Ethernet, and Zigbee can all be used by the HAN to communicate. Commercial Area Networks (CANs) and residential premises both use HANs, while industrial locations and commercial offices use Industrial Area Networks (IANs) [7].

A WANs connects smart meters, service providers, and electric utilities, enabling communication and data exchange across a vast geographical area. WANs can leverage multiple communication technologies, such as fiber optics, 3G/GSM/LTE, and WiMAX, to ensure reliable and efficient data transmission. To deliver the required data, the smart meter serves as a conduit between internal devices and outside parties. The electric utility manages power distribution within the smart grid. It also gathers detailed, sub-hourly power consumption data from smart meters and sends timely notifications to relevant parties. Devices within the HAN send messages to the smart meter, which forwards them to the relevant service provider [7] accordingly.

The reliance on Internet connectivity, cloud computing, and wireless technologies makes Smart Grids susceptible to Transmission Control Protocol/Internet Protocol (TCP/IP) based cyberattacks, such as hacking and data breaches (see Fig. 1). If security measures are not appropriately implemented, this could compromise the Confidentiality, Integrity, and Availability (CIA) of the Smart Grid [8].

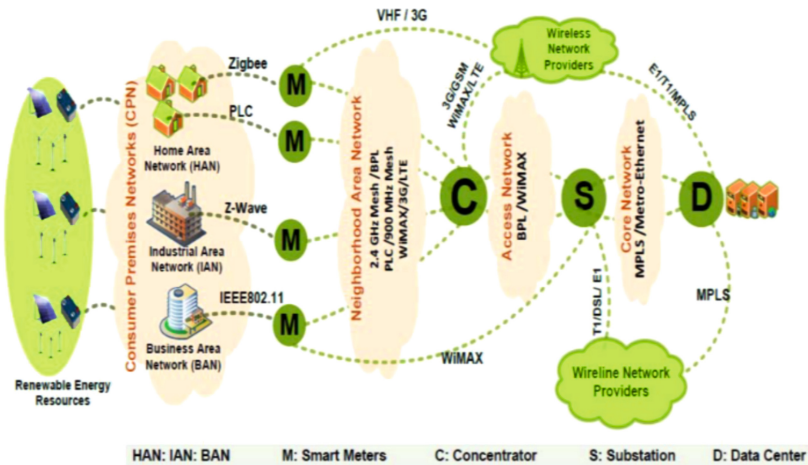


Fig. 1. Smart Grid basic network architecture [7].

1.2 Importance of Security SCADA in Smart Grid

By integrating SCADA systems, Smart Grids gain real-time visibility into their entire network, enabling remote monitoring and control of plant and equipment, thereby significantly enhancing their capabilities. Without SCADA, many critical Smart Grid operations would be significantly hindered, making it a vital component of the system [10–14]. Control systems are computer-based systems that underpin the operation of many critical infrastructures, enabling the monitoring and control of sensitive processes and physical functions [15–19]. A secure SCADA system is a prerequisite for deploying an intelligent Smart Grid, ensuring system stability, reliability, and automated control and monitoring [20].

The power system automation system, leveraging SCADA, provides a suite of functionalities including power control, contingency-based load shedding, and enhanced customer empowerment through integration with energy management systems [21–23]. Furthermore, SCADA contributes to the grid's self-healing capabilities by enabling autonomous responses to disturbances and system anomalies [24–27].

1.3 Contribution of the Study

The study's contribution is as follows: 1) The suggested ECS algorithm makes sure that the confidentiality, integrity, and availability cybersecurity goals for the Smart Grid are appropriately implemented and met. 2) By increasing the effectiveness of the encryption and decryption approaches, the integration of the N-th degree Truncated polynomial Ring Units (NTRU) algorithm with the Toom-Cook algorithm guaranteed that the suggested ECS algorithm provided secure end-to-end authentication and decreased complexity. 3) By quickly and effectively identifying harmful activity within the network, the proposed ECS algorithm stops cyber attackers from getting unauthorised access to the Smart Grid SCADA network, altering data, and launching modification attacks.

This article is organised as follows: it discusses the goals of smart grid cybersecurity and provides background information on SCADA and smart grid technologies. Section 2 of the document provides an overview of the literature and a list of recommended algorithms for safeguarding and securing SCADA systems in smart grids. The proposed algorithm is presented in Sect. 3 of this paper, which also describes how the Toom-Cook and NTRU algorithms are integrated. In Sect. 4, simulation Analyses and results are presented and discussed. Section 5 concludes the paper and presents the future work.

2 Related Work

The secure architecture and deployment of the Smart Grid have garnered significant attention. To safeguard the Smart Grid, numerous studies on different protocols, security algorithms, and key management systems have been carried out in recent years. The previously suggested fixes are covered in this section. This study report goes into further detail on the shortcomings of those suggested fixes.

To enhance the security of SCADA systems by ensuring data integrity, confidentiality, authentication, and non-repudiation, Premnath [2] introduced NTRU-based encryption and authentication systems. NTRU, a public-key encryption algorithm grounded in lattice-based cryptography, offers a robust solution to these security challenges. The various asymmetric methods were compared, and their encryption, decryption, and key-generation speeds were all evaluated in terms of performance. Comparative analysis reveals that NTRU cryptographic operations exhibit a performance advantage of 2–10 times over RSA and ECC for comparable security parameters. In addition, the hybrid solution's encryption and decryption times were compared to the times required for the RSA and NTRU digital signature algorithms. The outcomes demonstrated that NTRU could attain great speeds with little processing power. This shows that using an NTRU-style lightweight asymmetric key protocol is essential to enabling safe and quick real-time critical application support for SCADA systems. To evaluate the findings in a simulated SCADA device, the study must still combine and analyse the performances of the various cryptographic procedures that are currently in use in real-time SCADA protocols.

Das [4] proposed a cryptographic and key management framework aimed at securing data in light of the ongoing cybersecurity threats and issues facing the Smart Grid. The research presented several security algorithms designed to protect the Smart Grid and extensively examined various cryptographic and key management techniques. Among the symmetric algorithms highlighted were Triple Data Encryption Standard (TDES), Triple Data Encryption Algorithm (TDEA), and Advanced Encryption Standards (AES). On the other hand, the asymmetric algorithms showcased included Digital Signature Standard (DSS), Digital Signature Algorithm (DSA), and Elliptic Curve Digital Signature Algorithm (ECDSA). However, the study does not address consumer data privacy, which is also crucial for effectively safeguarding the Smart Grid and its stakeholders against potential cyberattacks. As a result, ensuring consumer data privacy remains a significant challenge.

Shahzad et al. [10] proposed a security solution for SCADA systems based on symmetric and asymmetric cryptographic algorithms, referred to as the SCADA security mechanism. This solution provides comprehensive end-to-end security for SCADA systems, protecting against communication attacks and detecting potential threats via the internet. To secure SCADA/DNP3 communications against authentication and confidentiality breaches, the Advanced Encryption Standard (AES) and Rivest-Shamir-Adleman (RSA) algorithms were utilized. Additionally, the SHA-2 hashing algorithm was implemented to safeguard SCADA/DNP3 from integrity attacks. To counter non-repudiation threats, a “digital signature” feature, based on SHA-2 hashing and the RSA method, was also introduced.

To evaluate the effectiveness of this proposed solution, a prototype of the Distributed Network Protocol 3 (DNP3) was developed, focusing on the “application layer, pseudo-transport layer, and data link layer” of the DNP3 protocol. A testbed configuration was set up, comprising seven Remote Terminal Units (RTUs) connected to Master Terminal Units (MTUs) at Stations 1 and 2. The seven RTUs were identified as iRTU1, RTU2, RTU3, RTU4, RTU5, RTU6, and RTU7. Various security threats, including data

injection, packet replay, brute force attacks, shared key guessing, key cracking, man-in-the-middle attacks, and data deletion, were simulated. According to the results of system behavior measurements, the proposed solution demonstrated strong security for authentication, data integrity, confidentiality, and non-repudiation when compared to alternative end-to-end security solutions. However, the limitation of Shahzad et al.'s solution [10] is that it is restricted to the DNP3 protocol, which is currently the only SCADA protocol in use. SCADA systems utilize multiple protocols, such as Profibus, Fieldbus, and Modbus, which were developed without security considerations. Therefore, there remains a need for the proposed security mechanism to be applied to other SCADA protocols currently in use.

Rezai et al. [16] conducted a survey examining the current key management techniques utilized to secure SCADA networks. They introduced two categories of key management methods: symmetric and asymmetric. One of the evaluation criteria for these techniques was the number of keys that need to be stored, which reflects the associated storage cost. Additionally, the overall delay time was analyzed, encompassing the durations involved in group key establishment, certificate verification, data transfer, and the processes of message encryption and decryption.

The findings indicated that effectively securing a SCADA network necessitates more than merely employing symmetric or asymmetric key management techniques. Certain Remote Terminal Units (RTUs) and Intelligent Electronic Devices (IEDs) are not compatible with Public Key Cryptography (PKC), indicating that relying solely on asymmetric key management is insufficient. Therefore, extensive research is needed to develop effective hybrid architectures that combine both asymmetric and symmetric key management approaches.

Moreover, Rezai et al. [16] highlighted several areas requiring further investigation, including public key infrastructures, management of cryptographic authorities and certificates within SCADA key management frameworks, identifying vulnerabilities in SCADA networks, implementing distributed security mechanisms that can operate within resource limitations, exploring cloud security applications in SCADA environments, and developing lightweight cryptographic algorithms and protocols. They also pointed out the need for new hardware tailored for SCADA networks, such as firewalls, intrusion detection systems (IDSs), access control measures, and mechanisms for assessing protocol vulnerabilities.

Nitaj's study [17] confirms Premnath's findings, highlighting NTRU as one of the fastest public-key cryptosystems available, introduced in 1996. NTRU is a post-quantum cryptosystem, resistant to attacks from quantum computers. Unlike traditional public-key cryptosystems like RSA, ECC, and El Gamal, which rely on the difficulty of factoring large numbers and computing discrete logarithms, NTRU's security is based on different mathematical principles. As quantum computers can efficiently solve these problems, NTRU offers a more secure alternative.

Ghosh and Sampalli [20] proposed an ID-based key management architecture using pairing-based cryptography on elliptic curves. This approach offers efficient key establishment, secret key recovery, and eliminates the need for digital certificates. However, like traditional public-key cryptography, these methods are vulnerable to quantum

attacks, as noted by Castellanos [24]. While efficient, they may not be future-proof against quantum computing advancements.

Miri et al. [25] proposed the NTRU cryptosystem as a suitable solution for securing communications and data in Ad-Hoc Ultra-Wide Band Impulse Radio (UWB-IR) networks. Their study compared three public-key cryptosystems: RSA, ECC, and NTRU, based on key size, security level, and encryption/decryption speed. The results demonstrated that NTRU outperforms both RSA and ECC in terms of encryption and decryption speed for all security levels. NTRU offers several advantages, including shorter key sizes, lower processing power requirements, easier key generation, and faster encryption/decryption speeds. Additionally, NTRU is more energy-efficient and has a smaller footprint compared to traditional public-key cryptosystems.

To improve the computational efficiency of cryptographic systems, Mera et al. [26] introduced the Toom-Cook algorithm for key encoding and decoding. This technique is also applied to McEliece cryptosystems to address limitations in key size and transmission rate. The Toom-Cook algorithm outperforms the Karatsuba method, reducing the number of multiplications by a factor of 9/5. While the classical method requires $O(n^2)$ multiplications, the Toom-Cook algorithm reduces this to $O(n^{1.465})$. Mera [26] concludes that the Toom-Cook algorithm remains one of the most effective methods for multiplying large integers.

Linardopoulou [27] suggested a practical implementation of the NTRU algorithm, identifying suitable algorithms for a secure implementation. However, the study noted limitations in Java's implicit conversion capabilities, which could impact future implementations. To enhance security, future implementations should focus on clearing arrays and stored variables. Additionally, incorporating the Karatsuba or Toom-Cook algorithms could improve efficiency.

Feng et al. [28] proposed a hybrid encryption scheme to protect NILM data from attacks like data falsification, manipulation, and eavesdropping. This scheme combines AES, RSA, and HMAC-SHA1 algorithms. AES encrypts user data, while RSA secures the AES key. HMAC-SHA1 ensures data integrity and sender authentication. A test platform was established to evaluate the performance and security of the proposed scheme, demonstrating its effectiveness in meeting data security and confidentiality requirements. However, the study did not address availability, a critical aspect of security.

Alese et al. [29] compared the performance of Menezes Vanstone Elliptic Curve Encryption (MVEC), ECC, and RSA encryption algorithms on an Intel Pentium dual-core system. The study evaluated key generation, encryption, and decryption times for various key sizes. Results showed that ECC significantly outperforms RSA in terms of speed and efficiency for all key lengths. However, the study noted that ECC's relative newness and the ongoing research into the Elliptic Curve Discrete Logarithm Problem (ECDLP) raise concerns about its long-term security compared to the more established RSA algorithm.

The next section presents the proposed algorithm developed to address the limitations of the existing solutions to improve the security of the smart grid SCADA System.

3 Proposed Algorithm

This paper's primary goal was to create a better cybersecurity algorithm that will fortify and enhance cybersecurity for smart grids. Three (3) goals must be met by cybersecurity for smart grid systems: availability, integrity, and confidentiality. Technology relies heavily on algorithms to provide safe encryption, authentication, and integrity. Encryption, decryption, authentication, and key management are essential for the effective deployment of cybersecurity in smart grid systems. The cybersecurity algorithm was created in a way that guarantees that malevolent users cannot access the Smart Grid network without authorisation and that the cybersecurity goals of data availability, confidentiality, and integrity are met. To achieve this the proposed ECAS was designed by integrating the NTRU algorithm and Toom-Cook algorithm. NTRU was chosen because it is among the quickest and most robust public cryptosystems, resistant to both classical and quantum assaults [2]. The Toom-Cook algorithm was chosen because it is among the finest methods for multiplying large integers, which in turn increases the efficiency of the encryption and decryption process by lowering the complexity involved in key encoding and decoding as well as the transmission rate.

3.1 NTRU Algorithm

The paper suggests integrating a quicker and more lightweight NTRU asymmetric cryptosystem into SCADA systems. The goal of the NTRU algorithm was to drastically cut down on memory and processing overheads. The two algorithms that comprise the NTRU cryptosystem are NTRUSign, which is used for digital signatures, and NTRUEncrypt, which is used for encryption. Compared to RSA encryption, which does modulo exponentiation under specific rules, and ECC encryption, which performs scalar multiplication for points on an elliptic curve, NTRU encryption carries out encryption and decryption via polynomial operations that can be done at faster speeds. Consequently, the NTRU is a faster and lighter method when compared since it offers simple key generation, fast encryption and decryption, and very few computational resources. Nonetheless, the suggested approach integrated the NTRU and Toom-Cook algorithms and considered data encryption and decryption as well as user authentication. The efficient and safe encryption and decryption for the Smart Grid is ensured by this integration. The fact that NTRU is immune to quantum assaults and is classified as a post-quantum cryptosystem is another factor in the decision to use it. On the other hand, NTRUSign is useful for SCADA applications that have real-time limitations.

3.2 Toom-Cook Algorithm

Since the Toom-Cook algorithm is faster than both the Karatsuba and Toom-Cook-4-way approaches, it was selected to be merged with the NTRU. Additionally, a study by Mera et al. [26] showed that the Toom-Cook algorithm is a 9/5 improvement over the Karatsuba approaches, which reduce the number of multiplications from 9 to 5. Large integer multiplication improves efficiency in just a small number of methods. The well-known methods Karatsuba, Toom-Cook-3, and Toom-Cook-4 aid in increasing the effectiveness of cryptosystems. These methods are applied in cryptographic algorithms, like the Diffie

Hellman key exchange protocol, ElGamal, RSA, and elliptical curve cryptosystems, to lessen the complexity of key encoding and decoding. McEliece Cryptosystems also employs this technique to get around issues with transmission rate and encrypted key size. The research examined different cybersecurity algorithms and well-known methods for enhancing cryptosystem efficiency. Thus, by combining the NTRU and the Toom-Cook algorithm, the study suggested an integrated cybersecurity method for the Smart Grid SCADA system. An increase in processing speed was made possible by this integration. As a result, encryption and decryption methods became more effective, ensuring safe network connection.

3.3 Enhanced Cybersecurity Algorithm

Phase Initialisation, Phase Authentication, Phase Key Generation, Phase Encryption, Phase Decryption, and Phase Arithmetic Operation comprise the six (6) steps of the proposed algorithm's process.

The various steps start from the initiation phase, Step 1. In this phase the Client Sender A wishes to transmit data or command to Client Sender B. Steps 2, 3, 4, and 5 follow, in which authentication takes place using the NTRUSign algorithm for digital signatures. Digital signature schemes are common authentication mechanisms used in the SCADA. According to Premnath [2], digital signature schemes are a type of public-key encryption for identifying a sender and preventing data falsification when data is sent from a receiving machine/client to a machine/client. Using the transmitting client's private key, client sender A will generate a signature for the desired data to be sent. Subsequently, the requested data and the signature data are transmitted to Client Sender B. In order to determine whether the desired data has been fabricated, Client Sender B verifies the signature data using a public key that matches the private key of Client Sender A's device. Step 2 is the signing key; Step 3 is the verification key; Step 4 is the digital document; and Step 5 is a valid signature. These are the four steps in the authentication procedure. In Step 6, which is the key-generation phase, Client Sender A starts by selecting random polynomials and creating keys.

A. Initialization phase

Step 1: Client Sender A wishes to transmit data or command to Client Sender B.

Authentication Phase

The NTRUSign algorithm will be used for authentication. A digital signature scheme consists of:

- a set of (hashes of) digital documents D
- a set of signatures S
- a set of randomized elements R
- a set K of pairs K_{sign} and K_{verify} consisting of linked signing and verification keys.

A signing key is a map;

$$K_{\text{sign}}: D \times R \rightarrow S$$

and a verification key is a map.

$$K_{\text{verify}}: D \times R \rightarrow \{ \text{Yes}, \text{No} \}$$

Sign and verification keys satisfy.

$$K_{\text{verify}}(d, s) = \text{Yes} \Leftrightarrow s = K_{\text{sign}}(d, r) \text{ for some } r \in R$$

Public Parameters:

Dimension parameter N , odd primes p and q , and norm-bound $B = \lfloor \sqrt{2N/4} \rfloor$

Step 2: Signing Key

A pair of (f, g) , vectors with f random mod 3, and g random p

Step 3: Verification Key

$$h = f^{-1} * g \pmod{q}$$

Step 4: Digital Documents

A document (hash) is a pair of mod p vectors (s, t_p)

Step 5: Valid Signatures

A signature on (s, t_p) for the signature key h is a pair of vectors (s, t) satisfying:

- $t \equiv s * h \pmod{q}$, i.e., $(s, t) \in L_h$.
- $(s, t) \equiv (s, t_p) \pmod{p}$.
- $\|s\|_\infty$ and $\|t\|_\infty$ are both $\leq 1/2 q - B$.

NTRUSign Parameters

The following should be practical while providing good security:

$$N = 661$$

$$p = 3$$

$$q = 9829081$$

$$B = 1487$$

$$k = 10$$

With these parameters, we have:

Pro (Signature is Accepted) $\approx 45\%$

Key and Signature Size ≈ 15864 bits

Bit Security ≈ 192 to 256

Algorithm 1: NTRUSign – Signing Algorithm

This algorithm computes the NTRUSign signature on a document (sp, tp) using the signing key (f, g)

1. Choose a random r with $\|r\|_{\infty} \leq \lfloor q/2p-1/2 \rfloor$.
 2. Set $s_0 = sp + pr$.
 3. Set $t_0 = h * s_0 \pmod{q}$.
 4. Compute $a = g^{-1} * (tp - t_0) \pmod{p}$.
 5. Set $s = s_0 + a * f$ and $t = t_0 + a * g$.
 6. If $\|s\|_{\infty}$ or $\|t\|_{\infty}$ is $> 1/2 q - B$, then REJECT. Go to Step 1
 7. Return (s, t).
-

The information flow between two (2) SCADA devices is depicted in Fig. 2, along with the application of the NTRUSign method to verify the transmitter and receiver's network communication. This is to ensure data availability and confidentiality by Smart Grid's cybersecurity criteria.

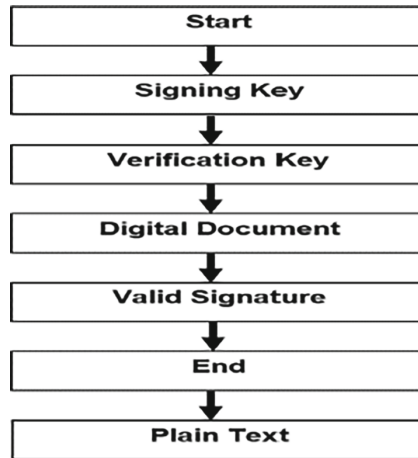


Fig. 2. Authentication process.

NTRU Key Generation algorithm is presented in Algorithm 2.

Algorithm 2: NTRU Key Generation

Output: $pk = (h)$, $sk = (f, fp, hq)$

8. $f, g \leftarrow \text{Sample}()$
 9. $f_q \leftarrow f^{-1} \bmod (q, \Phi_n)$
 10. $h \leftarrow (3 \cdot g \cdot f_q) \bmod (q, \Phi_1 \cdot \Phi_n)$
 11. $hq \leftarrow h^{-1} \bmod (q, \Phi_n)$
 12. $fp \leftarrow f^{-1} \bmod (3, \Phi_n)$
-

4 Simulation Results and Analysis

The NS-2 simulator was used to examine and report the performance findings to assess how well the integrated cybersecurity for the Smart Grid SCADA system was working. The network's behaviour was examined using the NAM. In addition, the NAM offered a visual representation of the nodes' positions, the packet transfers occurring on them, and other simulation specifics. The simulation's final findings were handled, analysed, and visually presented using XGraph. Ten simulations yielded the simulation results. Public key cryptosystems are compared and assessed based on their performance in terms of key size, security level, encryption and decryption speed, and packet loss:

- a) **Key size:** Key length is a crucial aspect of security. The number of bits in a key that a cryptographic method uses to encrypt a message is referred to as key length. The Smart Grid's encryption and authentication depend on key management. Since the length of the key defines the level of safety, it is imperative to both encode and decode the key. Consequently, increased security results from longer key sizes [33, 34].
- b) **Security level:** This is the degree of strength attained by a cryptographic primitive, like a hash function or cypher. Typically, the security degree is stated in terms of several "bits of security" (also known as security strength). For example, n -bit security indicates that it would take an attacker 2^n operations to breach it [34].
- c) **and decryption operation speed:** This refers to how long it takes to convert plain text into encrypted text. The time it takes to convert a decrypted message into legible text is measured by the decryption speed [34].
- d) **Packet Loss:** This happens when information-transporting data packets are sent over a network but are unable to arrive at their intended location [34].

The suggested algorithm, NTRU, was used in the study to execute, assess, and compare the cryptographic algorithms RSA and ECC. The study selected cryptographic algorithms because they are hard for adversaries or intruders to break due to their assumptions about computational difficulty [31–33]. Also referred to as PKCSs, RSA, and ECC are recognised methods for Smart Grid SCADA system security. Three algorithms are thought to be included in cryptosystems, which are used for key creation, encryption,

and decryption [2]. Both public and private keys are used in the encryption process. Data is encrypted using the public key (pk), which is known to all. Only the owner of the matching private key (sk) can decrypt the data, though. Because these current cryptosystems require more bandwidth, processing power, and memory, applying them in the Smart Grid for authorisation, authentication, encryption, and decryption presents a difficulty. Therefore, this study suggested using the NTRU cryptographic technique, which is quicker and lighter, to get around this problem.

4.1 Effects of Key Size and Security Level

A bit string used to change data is called a key. One crucial security factor in cryptographic methods is the key's bit size. The public-key algorithm sizes for the RSA, ECC, and NTRU algorithms are displayed in Table 1 along with corresponding security levels [31–34].

Table 1. Key sizes of Cryptographic Algorithms in Bits [34].

Security Level	n	RSA key sizes (bits)	ECC Key sizes (bits)	NTRU key sizes (bits)	Protection Life
80	251	1024	160	2008	Until 2010
112	347	2048	224	3033	Until 2030
128	397	3072	256	3501	Beyond 2030
192	587	7680	384	5193	Beyond 2030
256	787	15360	512	7690	--

The public-key size of a cryptosystem gives valuable insight into the bandwidth usage if the cryptosystem is intended to be used in key-exchange schemes [30]. Table 1 gives corresponding NTRU, ECC, and RSA key sizes for equivalent security levels (k) of 80 bits, 112 bits, 128 bits, 192 bits, and 256 bits.

The shortest vector problem (SVP), an extremely challenging lattice reduction issue, is the basis for the security of the NTRU, whereas the security of the RSA and ECC is based on the difficulty of factoring big composite integers or computing discrete logarithms. Because these algorithms were created based on the notion of computational hardness, it would be challenging for an opponent to undermine them in real life. Therefore, a quantum computer will be able to easily break RSA and ECC. In contrast, the NTRU is impervious to attacks utilising quantum computers.

The simulation's findings demonstrate that the RSA key size grows more quickly than the ECC key size as the symmetric key size increases (see Fig. 3). In comparison to RSA and NTRU, the ECC method can provide more security with each bit increase in key sizes. In terms of bandwidth efficiency, the ECC is the best option because it has the shortest key size. The more testing that is done, the more efficient the NTRU's bandwidth utilisation is in comparison to RSA.

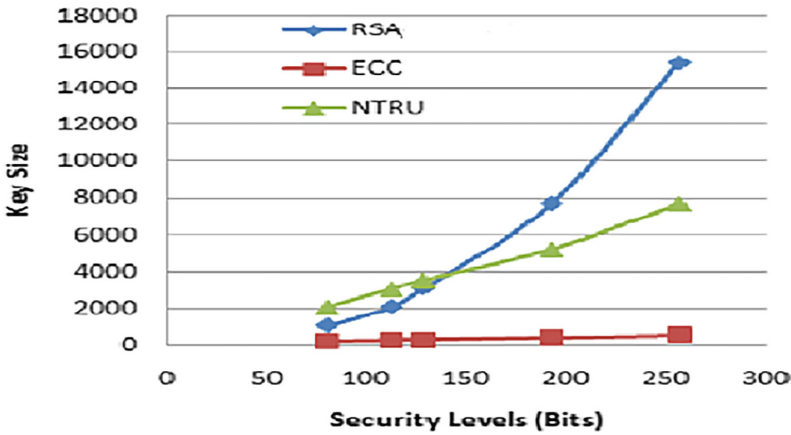


Fig. 3. Security level vs key size.

4.2 Performance Comparison of the RSA, ECC, and ECS Algorithms in Terms of Key Generation, Encryption, and Decryption

Table 2. Comparison of Key Generation, Encryption and Decryption Speed of RSA, ECC, and ECS Algorithms [34].

	RSA- 1024	ECC-168	ECS - 263
Public-key size (bits)	1024	168	1841
Key generation (ms)	1432	65	17.6
Encryption (ms)	4.32	140	1.4
Decryption (ms)	48.5	67	2.9

Key length is an important security parameter. Key length refers to the number of bits in a key used by a cryptographic algorithm to encrypt a message. Table 2 provides the corresponding NTRU, ECC, and RSA key sizes.

We tracked and compared the three (3) algorithms’ respective key generation speeds. We can draw the conclusion that compared to the ECC and RSA, the ECS algorithm’s performance gives a faster key generation. The findings show that the ECS algorithm, with a key-generation speed of 19.8 ms, is roughly three (3) times faster (see Fig. 4). It has also been observed, nevertheless, that ECC is quicker than RSA. Premnath [2] states that the usage of algorithms like the RSA or ECC for communication security is mandated by existing security standards for Smart Grid SCADA System communication. However, when applied to the Smart Grid SCADA system, with its real-time limits and hardware limitations, there are some performance concerns with the cryptographic solutions of these specifications. This analysis and evaluation clearly show that the ECS uses less battery life and CPU power, which makes it appropriate for devices with

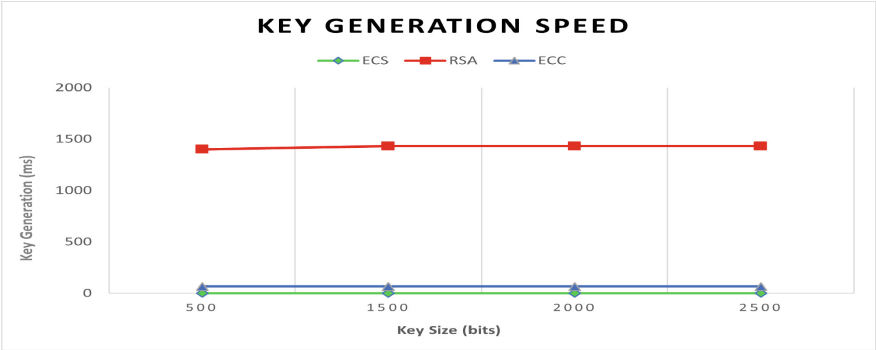


Fig. 4. Comparison of key-generation speed for RSA, ECC, and ECS algorithms.

limited resources when code size is a significant constraint. Because the ECS technique minimises computational and memory overheads, it can be applied to SCADA systems that have hardware limits and real-time constraints.

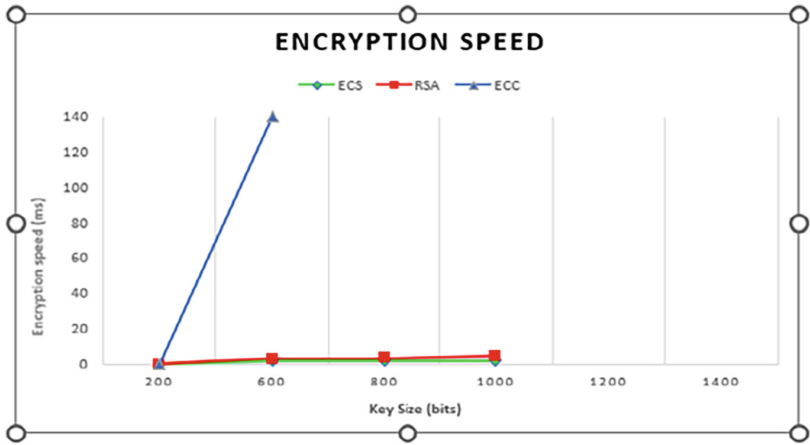


Fig. 5. Comparison of encryption speed for RSA, ECC, and ECS algorithm.

Data is transformed through encryption into a format known as a cypher text, which makes it difficult for unauthorised parties to decipher. The key component of any Smart Grid SCADA system’s data confidentiality and integrity protection is the encryption scheme’s design and selection [2]. A 32-byte message that was created at random was used for the test. The NTRU outperforms the ECC in terms of performance. According to the findings, ECS block encryption outperforms ECC block encryption by over 73 times. The ECS takes 1.4 ms to encrypt data, while the ECC takes 140 ms and the RSA takes 4.32 ms (Fig. 5). In this instance, RSA performs better than the ECC technique in terms of key-generation speed. ECS makes use of NTRUEncrypt, which uses polynomial operations for both encryption and decryption. Implementing such can happen more quickly than with RSA encryption. Under specific conditions, the NTRUEncrypt conducts modulo

exponentiation, and the ECC multiplies scalar values at locations on an elliptic curve [2]. The test results unequivocally demonstrate that, at the same security level, the ECS algorithm encryption achieves faster processing than the RSA and ECC algorithms.

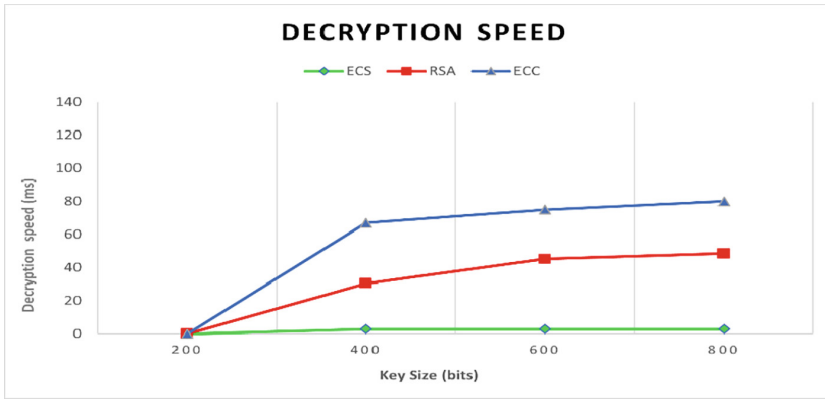


Fig. 6. Comparison of decryption speed for RSA, ECC, and ECS algorithm.

Cryptographic techniques such as encryption and decryption are used to produce secure information and communication. The process of restoring encrypted material to its original format so that the intended receiver can understand it is known as decryption. The findings for decryption speed are shown in Fig. 6. The duration of the algorithm's key decryption process is determined by the decryption speed parameter. A 32-byte message that was created at random was used for the test. The overall time required for ECS decryption, according to observed data, is 2.9 ms. The NTRUEncrypt, the strongest, lightest-weight, and quickest public-key cryptosystem currently in use, is used by the ECS algorithm. On the other hand, the ECC came in at 67 ms and the RSA at 48.5 ms. Due to the modular exponentiation of large integers involved in the RSA decryption process, it is costly.

4.3 Performance Comparison of the RSA, ECC, and ECS Algorithms in Terms of the Amount of Packet Loss

The parameter calculates the likelihood of packet loss while data is being transmitted and when MTUs and RTUs are communicating. After starting the conversation, the MTU sends the RTU data (bytes). Data in the form of random bytes has been sent many times between the MTU and RTU. Using a bandwidth of up to 5 Mbps, the data/information has been delivered and received multiple times between the MTU and RTU. As a result, the researcher can keep an eye on and compile data regarding the amount of packet loss that occurs during data transfer and communication.

Several runs of the experiment were conducted, and performance outcomes in both typical and atypical communication and traffic were noted. According to the test findings, 1.8% of packets were lost by both RSA and ECC, however the suggested ECS technique

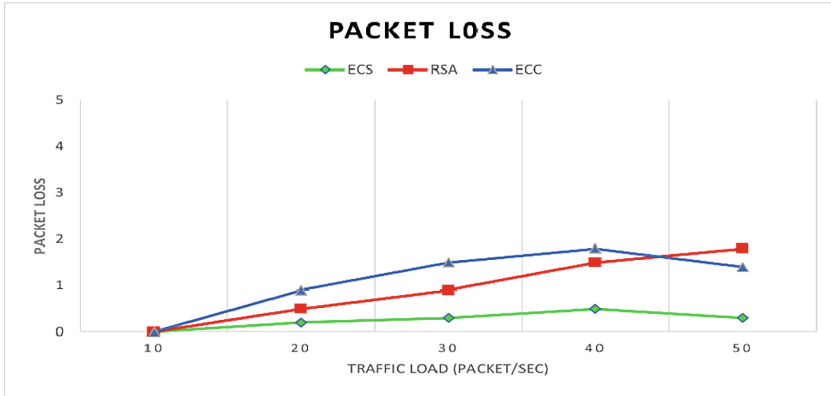


Fig. 7. Comparison of packet loss for RSA, ECC, and ECS algorithms.

only lost a maximum of 0.5% of packets during a DoS attack (see Fig. 7). A denial-of-service (DoS) attack aims to interfere with certain authorised operations, including stopping RTUs from corresponding with one another. When a DoS attack affects the RTU, orders sent from the HMI or MTU might not get through to the RTU. Unauthorised modifications to programmed instructions or status values in RTUs can cause equipment damage, cause operations to end early, or even disable control equipment, which could trigger a blackout. The performance assessment unequivocally demonstrates an improvement in availability attacks. For the Smart Grid SCADA system, availability and cybersecurity objectives come first. DoS targeting availability. As a result, packet loss becomes an essential testing parameter to make sure that the MTU's intended programming instructions and status reach the RTUs without being interrupted maliciously.

5 Conclusion and Future Work

The widespread occurrence of cyberattacks directed at crucial infrastructure is demonstrated by the 2019 and 2015 assaults on City Power in South Africa and the electrical system in Ukraine. The instances highlight the pressing necessity for improved cybersecurity protocols within the framework of smart grids. SCADA systems are especially susceptible to assaults because of their intricate and widely distributed design as well as their reliance on networked equipment. Although the NTRU algorithm was effectively included in the Toom-Cook algorithm for improved security in this paper, it is important to recognise the study's shortcomings. Subsequent investigations ought to concentrate on assessing the suggested ECS algorithm in diverse network circumstances, encompassing varying traffic volumes and assault situations. To guarantee that the ECS method is practically applicable in actual Smart Grid scenarios, it is also essential to examine the effect it has on resource usage and network performance. Additionally, examining the possible weaknesses and defences related to NTRUSign authentication would offer important perspectives on how secure it is.

Acknowledgment. The Tshwane University of Technology is acknowledged by the authors for its financial assistance. The authors affirm that there is no conflict of interest associated with this paper's publication.

References

1. Alsaleem, M.A., Yaseen, D.S., Aziz, S.: Efficiency comparison of Network Simulators. *Int. J. Adv. Sci. Technol.* **28**(6), 45–50 (2019)
2. Premnath, A.P.: Application of NTRU Cryptographic Algorithm for securing SCADA communication. UNLV Theses, Dissertations, Professional Papers, and Capstones, pp. 1–54 (2018)
3. Chhimwal, P., Rai, D.S., Rawat, D.: Comparison between different wireless sensor simulation tools. *IOSR J. Electron. Commun. Eng. (IOSR-JECE)* **5**(2), 54–60 (2013)
4. Das, A.: Cryptographic and key management approach for cybersecurity in smart grid. *Int. J. Eng. Res. Technol.* **4**(6), 63–72 (2021)
5. Feng, R., et al.: A Hybrid cryptography scheme for NILM data security. *Electronics* **9**(11), 1–18 (2020)
6. Kabir, M.H., Islam, S., Hossain, J., Hossain, S.: Detail comparison of network simulators. *Int. J. Sci. Eng. Res.* **5**(10), 204–218 (2014)
7. Saidallah, M., Fergougui, A.E.L., Elalaoui, A.E.: A survey and comparative study of open-source wireless sensor network simulators. *Int. J. Adv. Res. Comput. Sci.* **7**(3), 1–7 (2017)
8. Sayed, K. and Gabbar, H.A.: SCADA and smart energy grid control automation. In: *Book: Smart Energy Grid Engineering*, pp. 481–514. Academic Press (2017)
9. Seemma, P.S., Nandhini, S., Sowmiya, M.: Overview of cyber security. *Int. J. Adv. Res. Comput. Commun. Eng.* **7**(11), 125–128 (2018)
10. Shahzad, A., Musa, S., Irfan, M.: N-secure cryptography solution for SCADA security enhancement. *Trends Appl. Sci. Res.* **9**(1), 381–395 (2014)
11. Sharma, R., Vashisht, V., Singh, U.: Modelling and simulation frameworks for wireless sensor networks: a comparative study. *Inst. Eng. Technol.* **10**(5), 181–197 (2020)
12. Vijayapriya, T., Kothari, D.P.: Smart grid: an overview. *Scientific Research on Smart Grid and Renewable Energy (SGRE)*, pp. 306–310 (2020)
13. Wang, Y.: Smart grid, automation, and SCADA systems security. *World Scientific Review*, pp. 1–16 (2012)
14. Wu, J. and Kobara, K.: Comparison of tools and simulators for control system security studies. *Research Institute for Secure System (RISEC) and National Institute of Advanced Industrial Science and Technology (AIST)*, pp. 1–6 (2012)
15. Yadav, G., Paul, K.: Assessment of SCADA system vulnerabilities. *IEEE Access* **2**(3), 1737–1744 (2019)
16. Rezai, A., Parviz., Keshavarzi, P., Moravej, A.: Key management issue in SCADA networks: a review. *Eng. Sci. Technol. Int. J.* **20**(1), 354–362 (2017)
17. Nitaj, A.: The mathematics of the NTRU public key cryptosystem, pp. 1–16 (2015)
18. CyberEdge Group: Cyberthreat Défense Report, pp. 1–66 (2021)
19. Amin, B.M.R., Taghizadeh, S., Rahman, S., Hossain, J., Varadharajan, V., Chen, Z.: Cyber attacks in smart grid – dynamic impacts, analyses and recommendations. *Inst. Eng. Technol.* **5**(4), 321–329 (2020)
20. Ghosh, S., Sampalli, S.: Survey of security in SCADA networks: current issues and future challenges. *IEEE Access* **7**(1), 135812–135831 (2019)
21. NISTIR 7628 Revision 1: Guidelines for Smart Grid Cybersecurity. Volume 1 - Smart Grid Cybersecurity Strategy, Architecture, and High-Level Requirements, pp. 1–66 (2014)

22. Latest Hacking News. <https://latesthackingnews.com/2020/02/17/south-african-nedbank-suffered-data-breach-affecting-1-7-million-customers/>. Accessed 25 July 2020
23. Aloula, F., Al-Alia, A.R., Al-Dalkya, R., Al-Mardinia, M., El-Hajjb, W.: Smart grid cyber security: challenges and solutions. *Int. J. Smart Grid Clean Energy* **2**(3), 1–7 (2016)
24. Castellanos, S.: Nascent quantum computing poses threat to cybersecurity. *CIO J.- WSJ* **1**(1), 1–10 (2017)
25. Miri, J., Nsiri, B., Bouallegue, R.: Certificateless scheme based NTRU cryptosystem for ad-hoc UWB-IR network. *Int. J. Wireless Mobile Netw.* **9**(6), 53–69 (2017)
26. Mera, J.M.B., Karmakar, A., Verbauwhede, I.: Time-memory trade-off in Toom-Cook multiplication: an application to module-lattice based cryptography, pp. 1–23. *The International Association for Cryptologic Research (IACR)* (2020)
27. Linardopoulou, T.: Implementation and analysis of the NTRU algorithm in java, pp. 1–33. *Bachelor of Science Computer Science & Systems University of Washington Tacoma* (2020)
28. Feng, R., et al.: A hybrid cryptography scheme for film data security. *Electronics* **9**(7), 1–28 (2020)
29. Alese, B.K., Philemon, E.D., Falaki, S.O.: Comparative analysis of public-key encryption schemes. *Int. J. Eng. Technol.* **2**(9), 1552–1568 (2012)
30. Loriya, H., Kulshreshta A., Keraliya, D.: Security analysis of various public key cryptosystems for authentication and key agreement in wireless communication network. *Int. J. Adv. Res. Comput. Commun. Eng.* **6**(2), 267–274 (2017)
31. Miri, J., Nsiri, B., Bouallegue, R.: Certificateless scheme based NTRU cryptosystem for Ad-Hoc UWB-IR network. *Int. J. Wireless Mobile Networks* **9**(6), 53–69 (2017)
32. Kabir, M.H., Islam, S., Hossain, M.J., Hossain, S.: Detail comparison of network simulators. *Int. J. Sci. Eng. Res.* **5**(10), 203–218 (2014)
33. Abood, O.G., Guirguis, S.K.: A survey on cryptography algorithms. *Int. J. Sci. Res. Publ.* **8**(7), 495–516 (2018)
34. Halefose, K.H., Mathonsi, T.E., Muchenje, T., Du Plessis, D.P.: Implementation of an enhanced cybersecurity algorithm for smart grid. In *2022 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 1028–1033 (2022)



Improving Face Recognition from a Single Image

Justice Owusu Agyemang^(✉), Desmond Delali Atakpla,
Awumee Gabriel Selorm, Thomas Boadu, Martin Kodua-Basoa,
and Christabel Bemah Darko

Faculty of Electrical/Computer Engineering, Kwame Nkrumah University of Science
and Technology, Kumasi, Ghana

`jay@knust.edu.gh,`

`{ddatakpla1,sgawumee,tboadu,mkbasoa,cbdarko}@st.knust.edu.gh`

Abstract. Facial recognition technology has become an essential tool for various security applications. However, traditional face recognition systems require several images of an individual to be effective. Training systems with only a single image per person remains a significant challenge. This paper presents a face-recognition approach using a single image by leveraging deep learning techniques. We propose a method that generates multiple variations of a single image through image segmentation and preprocessing techniques, such as de-noising and illumination correction. The core of our approach involves using Siamese neural networks, which are trained to distinguish between similar and dissimilar image pairs, enabling accurate recognition even with minimal input data. Our methodology includes developing and training a Siamese neural network focusing on metric-based learning for one-shot face recognition. We evaluate our approach on the Celebrity dataset, achieving 97.5% accuracy and demonstrating competitive performance compared to existing single-image face recognition methods, even with limited input data.

Keywords: Facial Recognition · Siamese Neural Network (SNN) · Deep Learning · One-shot Learning

1 Introduction

Face recognition has garnered significant attention over the past decades due to its growing demand in security applications, such as video and biometric surveillance. It is one of the most important biometric methods. While other techniques like fingerprint and iris recognition are often more reliable, they tend to be intrusive and heavily reliant on user cooperation. Face recognition, on the other hand, stands out as a universal, non-intrusive, and easily accessible system. It is user-friendly and efficient for mass scanning, a challenge for other biometric methods. Moreover, face recognition is natural, socially accepted, and avoids the health risks associated with shared equipment that might transmit germs or impurities.

Modern facilities such as hospitals, airports, and banks are increasingly integrating face recognition technology into their security systems. Despite its current success, research in this field continues to focus on improving speed and accuracy. The performance of any face recognition system is strongly dependent on its underlying face detection system; a robust face detection system enhances the overall effectiveness of the recognition process. Typically, face detection involves identifying human faces from images or live video feeds with human presence.

Most facial recognition systems require multiple images of an individual for training to achieve reliable recognition. However, in scenarios where only a single image is available, traditional systems face limitations. This raises the question: can a system generate variations of a single image to simulate the multiple images needed for training? Neural networks can be employed to address this challenge by generating augmented versions of the single image with variations such as lighting intensity and segmentation. Machine learning techniques can then process these augmented images to train the recognition system effectively. However, challenges remain, such as distinguishing between individuals with identical facial features, like twins, where recognition errors may occur.

The primary objective of this study is to develop a mechanism for training face recognition systems using a single image, leveraging neural networks. To achieve this, the study focuses on identifying suitable image segmentation techniques to generate multiple variations from a single image, applying pre-processing techniques such as de-noising and illumination correction, and evaluating the performance of deep learning algorithms on single-image-based face recognition systems.

The structure of this paper is organized as follows: Sect. 2 reviews related work and outlines the objectives. Section 3 discusses the proposed methods. Section 4 presents the results of the method evaluations. Finally, Sect. 5 concludes the paper and provides recommendations for future research.

2 Related Works

This section provides an overview of face recognition systems, including a discussion of related research and their operational principles. Among the various techniques available for object recognition, deep learning has emerged as a widely adopted approach.

Fadhlan et al. [1] introduced an appearance-based method tailored for single sample per person (SSPP) face recognition. Their approach involved holistic processing of face images using Principal Component Analysis (PCA) combined with eigenfaces, complemented by leveraging biased knowledge obtained from a trained artificial neural network (ANN). In this method, biased information was used to distinguish intra-subject variations from inter-subject differences, thereby narrowing down the set of possible matches to a single individual. Subsequently, the remaining candidates were prioritized based on a confidence score generated by PCA and eigenface techniques, facilitating the selection of the

most likely match from the SSPP gallery. This candidate-ranking strategy, which aligns with the recommendations outlined in [2], is depicted in Fig. 1.

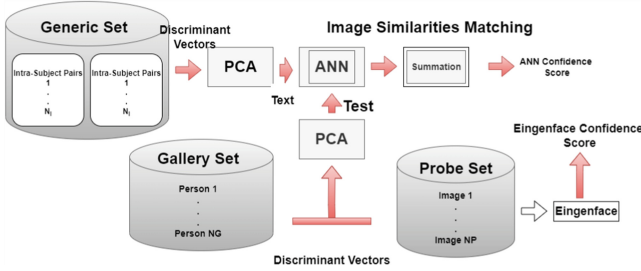


Fig. 1. The framework of Face Recognition From Single Sample Per Person by Learning of Generic Discriminant Vectors.

Junying et al. [3] proposed a framework combining Traditional Learning (TL) and Deep Learning (DL) approaches, known as TDL, as shown in Fig. 2. The TDL framework incorporates transfer learning to address the challenges of Single Sample Per Person (SSPP) face recognition, eliminating the need for extensive datasets to train deep convolutional neural networks (DCNNs). The framework involves a sample expansion method to overcome limited data in SSPP scenarios, fine-tuning a pre-trained DCNN model using selected samples from the expanded dataset, and evaluating the fine-tuned model for its performance. Additionally, the authors developed an intraclass variation set to further enrich the dataset by generating additional facial samples, enhancing the model's robustness.

Fenglian Li et al. [4] proposed a face recognition method that combines virtual sample generation with multi-scale feature extraction to improve performance. To augment the training dataset, they introduced the NMF-MSB technique, which integrates Non-negative Matrix Factorization (NMF) with three extensions: Mirror transform (M), Sliding window (S), and Bit plane (B). For feature extraction, they developed the WPD-HOG-P approach, which utilizes Wavelet Packet Decomposition (WPD), Histograms of Oriented Gradients (HOG), and an Image Pyramid to capture multi-scale facial features. These extracted features were then used to train a recognition model, optimized with a grid search-based support vector machine (SVM). Performance testing on the ORL and FERET datasets showed that this method outperformed traditional techniques in terms of recognition accuracy and computational efficiency. The flowchart illustrating the proposed NMF-MSB sample generation method is shown in Fig. 3.

Thus, for a single sample per person, nine virtual samples A1 to A9 were generated. The final training samples were composed of the original single sample and the nine virtual samples generated. These samples provided more useful information for improving the recognition performance of SSPP.

Tsung-Han Tsai et al. [5] presented a multi-task training framework that leverages a feature pyramid and triplet loss for training a single-stage deep neural network capable of both face detection and recognition. By utilizing a shared

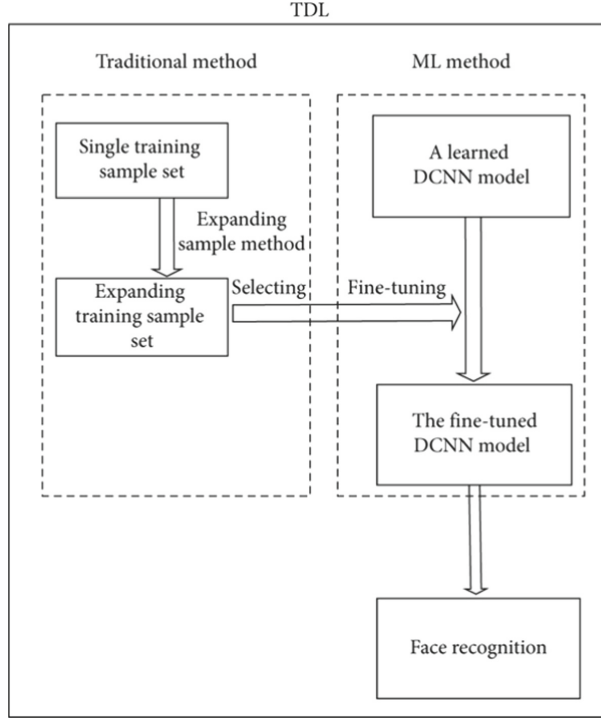


Fig. 2. The framework of the proposed method.

backbone network, the approach avoids redundant computation and optimizes memory usage by sharing weights across tasks. The network architecture (Fig. 4) integrates a feature pyramid and anchor boxes for face localization, employs triplet loss for feature extraction, and performs feature matching with a straightforward mathematical function. This method offers significant computational efficiency, achieving a processing speed of 212 FPS for a 640×640 resolution input on an Nvidia 2080Ti GPU while maintaining a recognition accuracy of 92.4% on the LFW dataset.

Vargas et al. [6] proposed a method for training Siamese neural networks with a novel structure designed to rank image similarity effectively. The proposed approach demonstrates a substantial improvement over existing baselines, delivering performance on par with state-of-the-art methods. Wu et al. [7] introduced a novel activation function named Max-Feature-Map (MFM) as part of their light-CNN model. This design reduces the number of parameters and enhances computational efficiency, allowing for robust learning of facial representations, even with noisy labeled datasets. In a similar vein, Zheng et al. [8] proposed a normalized light-CNN structure featuring 11 hidden layers, achieving a face verification accuracy of 98.46% on the Labeled Faces in the Wild (LFW) dataset. Furthermore, Schroff et al. [9] developed FaceNet, a system that projects facial

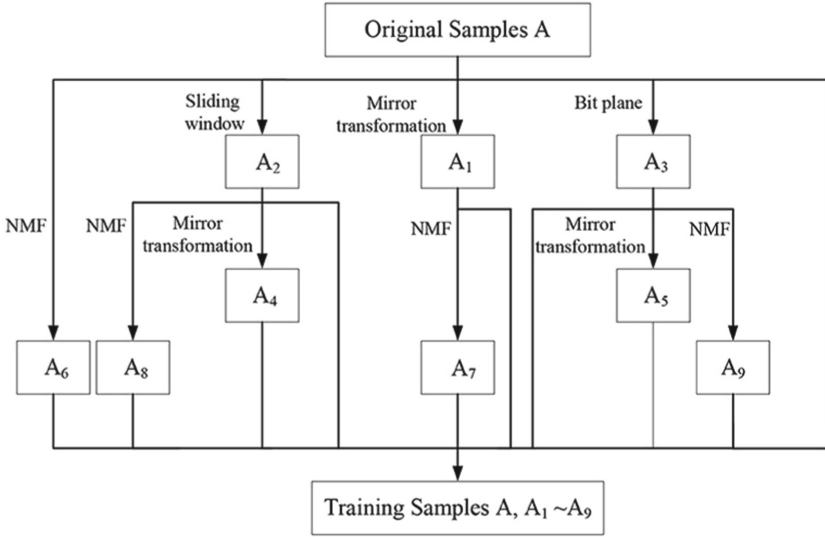


Fig. 3. The flowchart of the proposed NMF-Mirror transform (M), Sliding window (S), and Bit plane (B) sample generation method [4].

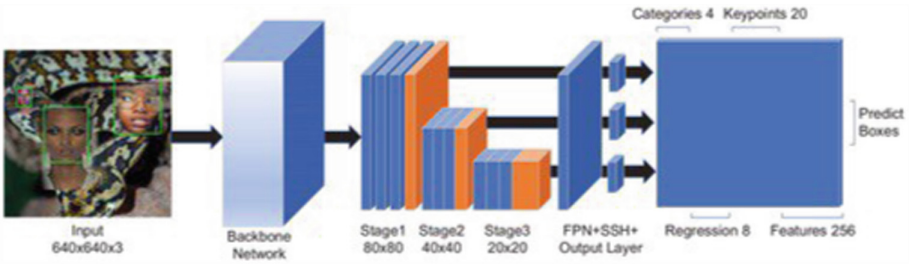


Fig. 4. The proposed model architecture [5].

images into Euclidean space to enable distance-based face recognition. Leveraging a deep convolutional network for feature extraction, FaceNet reached an accuracy of 99.63% on the LFW dataset and 95.12% on the YouTube Faces dataset.

One of the most critical stages in face recognition models is feature extraction. Hand-crafted features such as eigenfaces [10], Fisher-faces [11], and Local Binary Patterns (LBP) [12, 13] have demonstrated strong performance in face recognition tasks. However, their effectiveness diminishes in unconstrained environments characterized by complex backgrounds, varying illumination, diverse poses, and occlusion [12]. By leveraging multiple CNNs and modifying the loss function, Yu et al. [14] achieved higher accuracy compared to traditional approaches.

3 Methodology

This section outlines the step-by-step approach adopted to address the problems identified in the study. The general proposed architecture is depicted in Fig. 5.

This research employed a supervised metric-based methodology using Siamese neural networks to learn image representations. The features derived from the trained model were subsequently utilized for one-shot learning, eliminating the need for retraining.

Siamese Neural Networks (SNNs) are a specialized deep learning framework developed to assess the difference between two images. The architecture is referred to as “Siamese” because it comprises two identical subnetworks with shared weights. The outputs of these subnetworks are merged and passed through a fully connected layer to compute a similarity measure.

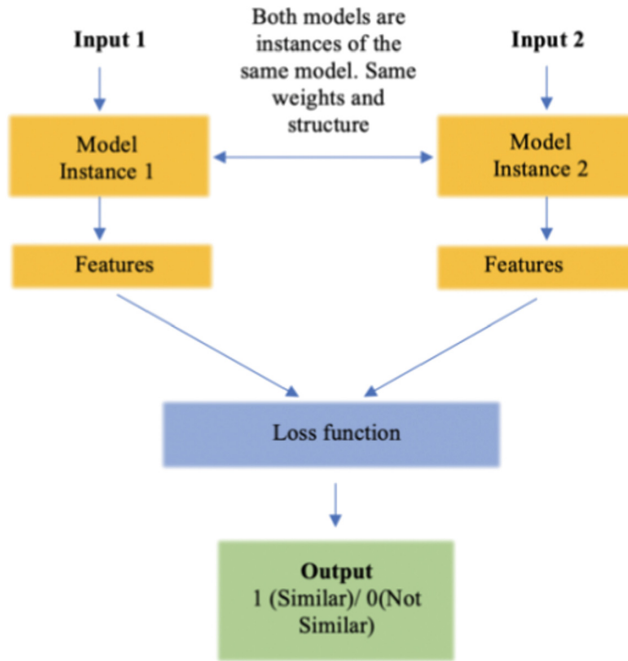


Fig. 5. Methodology

This research centers on image recognition, although the methodology can be extended to other data modalities. Specifically, we utilized large Siamese convolutional neural networks that:

- Extract versatile image features capable of supporting predictions for unseen class distributions, even when provided with minimal examples from new categories.

- Can be effectively trained with standard optimization algorithms on paired samples from the source dataset, providing a robust solution that bypasses the need for domain-specific knowledge while leveraging the strengths of deep learning.

To construct a model for one-shot image classification, we developed a neural network designed to differentiate between the class identities of image pairs, following a conventional verification task framework for image recognition. We posited that networks proficient in such verification tasks would naturally generalize to one-shot classification scenarios.

The verification model was trained to evaluate input pairs by estimating the probability that they belong to the same or different classes. This model was then employed for one-shot classification, where a novel class is represented by a single example. During testing, the model performed pairwise comparisons between a test image and each novel class example, assigning the highest probability to the class with the best matching pair.

3.1 Data Preparation

This study utilized the online Celebrity dataset, which contains images of 1,680 celebrities, with 2 to 10 images per individual. The dataset was processed for training a deep learning model using the triplet loss method. After loading and splitting the images into training and testing sets, triplets of the form (anchor, positive, negative) were generated. Here, the anchor and positive represent different images of the same celebrity, while the negative represents an image of a different celebrity.

In total, 9,164 triplets were created, with 7,331 triplets (80%) used for training and 1,833 triplets (20%) reserved for testing. This triplet-based approach enabled the model to learn discriminative features by minimizing the distance between anchor-positive pairs while maximizing the distance between anchor-negative pairs, providing a robust framework for facial recognition tasks.

3.2 The Model

The standard architecture utilized is a Siamese convolutional neural network comprising L layers, where each layer contains N_l units. The hidden vector in layer l for the first twin is represented by $\mathbf{h}_{1,l}$, while $\mathbf{h}_{2,l}$ denotes the equivalent for the second twin. Rectified Linear Units (ReLU) are employed in all layers except the final two, where sigmoidal units are used instead.

The network is structured as a series of convolutional layers, each utilizing a single channel with filters of adjustable sizes and a fixed stride of 1. The number of convolutional filters is chosen as a multiple of 16 to enhance computational efficiency. ReLU activations are applied to the resulting feature maps, optionally followed by max-pooling operations with a filter size and stride set to 2.

The k -th filter map at each layer is computed as follows:

$$a_{1,m}^{(k)} = \text{max-pool} \left(\max \left(0, \mathbf{W}_{l-1,1}^{(k)} * \mathbf{h}_{1,(l-1)} + b_l \right), 2 \right), \quad (1)$$

$$a_{2,m}^{(k)} = \text{max-pool} \left(\max \left(0, \mathbf{W}_{l-1,1}^{(k)} * \mathbf{h}_{2,(l-1)} + b_l \right), 2 \right), \quad (2)$$

where $\mathbf{W}_{l-1,l}^{(k)}$ represents the 3-dimensional tensor of convolutional filters for layer l , and $*$ denotes the “valid” convolution operation, which produces outputs only when the convolutional filter completely overlaps with the input feature map.

The output of the last convolutional layer is transformed into a flat vector and fed into a fully connected layer. This layer calculates the distance metric between the two branches of the Siamese network. The computed metric is then passed through a single sigmoidal unit to produce the final output.

More precisely, the prediction vector is given as:

$$\mathbf{p} = \sigma \left(\alpha_j \left| \mathbf{h}_1^{(j)} - \mathbf{h}_2^{(j)} \right| \right),$$

where $\mathbf{h}_1^{(j)}$ and $\mathbf{h}_2^{(j)}$ are the hidden representations of the two twins at the j -th layer, and σ denotes the sigmoid function (Fig. 6).

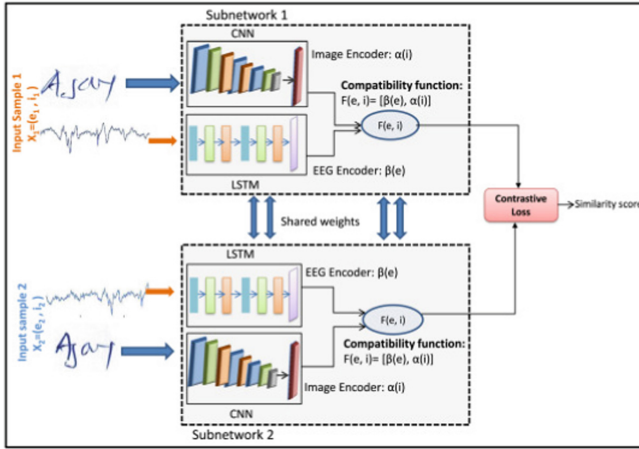


Fig. 6. Siamese Neural Network Architecture selected for verification task.

The largest version of the model used in this study provided the best results for the verification task. Figure 7 shows the design of a single encoder from the Siamese network.

3.3 Learning Loss Function

The Triplet Loss function, illustrated in Fig. 8, is designed to minimize the distance between the anchor (a) and the positive sample (p), while maximizing the

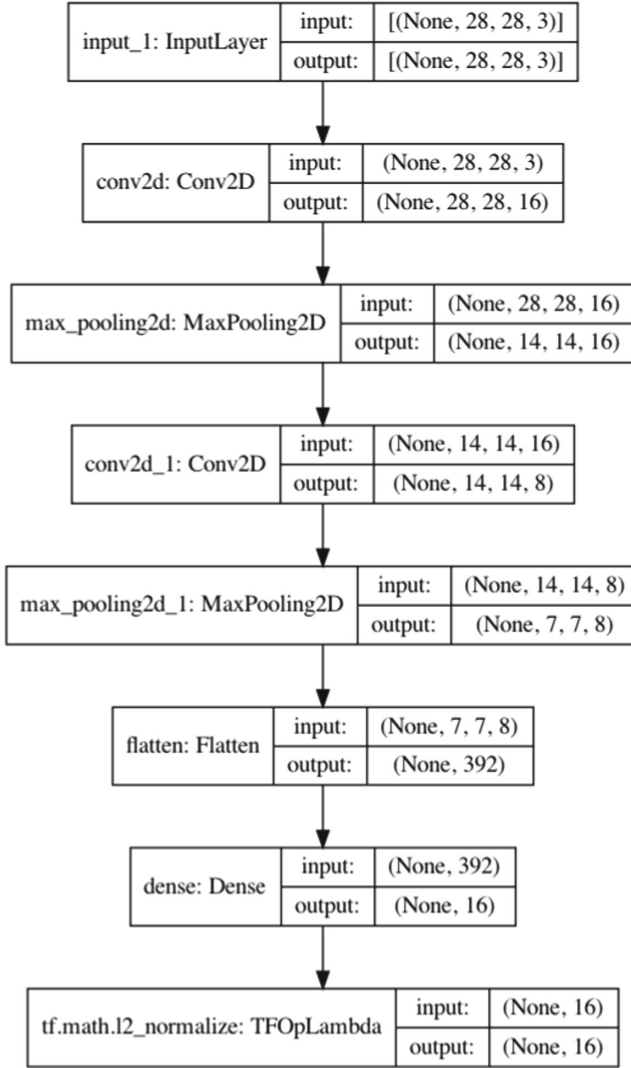


Fig. 7. A diagram of a single encoder from the Siamese network

distance between the anchor (a) and the negative sample (n). Mathematically, this is expressed as:

$$d(a, p) - d(a, n) < 0 \quad (3)$$

where d denotes a distance metric.

To ensure the loss is non-negative, we modify the equation as:

$$L = \max(d(a, n) - d(a, p), 0) \quad (4)$$

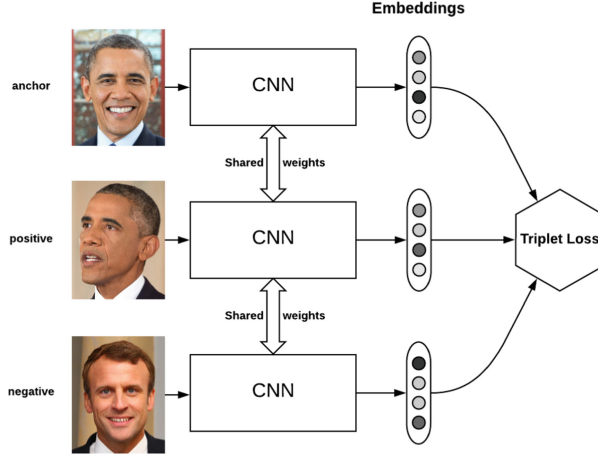


Fig. 8. Triplet Loss Function

To further increase the separation between positive and negative samples, a margin α is introduced, leading to the following Triplet Loss function:

$$L = \max(d(a, n) - d(a, p) + \alpha, 0) \quad (5)$$

This margin α ensures that the negative is at least α distance away from the positive.

We started by initializing the weights in the convolutional layers using a normal distribution with a mean of zero and a standard deviation of 10^{-2} . The biases were also initialized from a normal distribution, but with a mean of 0.5 and a standard deviation of 10^{-2} . This is represented as:

$$W_{conv} \sim \mathcal{N}(0, 10^{-2}) \quad (6)$$

$$b_{conv} \sim \mathcal{N}(0.5, 10^{-2}) \quad (7)$$

In the fully-connected layers, the biases were initialized in the same manner as the convolutional layers, but the weights were drawn from a wider normal distribution with zero mean and a standard deviation of 2×10^{-1} :

$$W_{fc} \sim \mathcal{N}(0, 2 \times 10^{-1})$$

$$b_{fc} \sim \mathcal{N}(0.5, 10^{-2}) \quad (8)$$

Although each layer had its own learning rate, the learning rate decayed uniformly by 1% per epoch across the network, as shown in the following equation:

$$\eta(T) = 0.99\eta(T - 1) \quad (9)$$

We observed that this learning rate annealing helped the network converge to local minima more efficiently and prevented it from getting stuck in unfavorable regions of the error surface. The momentum started at 0.5 for every layer and increased linearly over the epochs until it reached μ_j , the specific momentum term for the j -th layer. The networks were trained for a maximum of 500 epochs.

For testing the model, we loaded it and evaluated its performance on unseen images. Initially, we used a single encoder model to extract the features of an image, which were then plotted. These features, having a length of 16, were visualized using a scatter plot to examine their separation. Next, we used the output from the Siamese network (ranging from 0 to 1) to create a confusion matrix.

4 Results and Discussion

This section talks about the results obtained after undertaking the steps outlined in Sect. 2. The similarity measure takes the text-me data associated with a pair of images and returns a score between 0 and 1 on how similar the images are expected to be based on the text. Since there is a large number of pairs that can be generated even from a small dataset, the operation should not be prohibitively expensive as it runs many times, depending on the sampling.

For each epoch, we calculate and print the loss (which is used for optimization) and a few metrics to do with the validation. Over 500 iterations loss in the model reduces and spikes over the first 100 epochs and afterward reduces significantly and stays constantly over the rest of the iterations of the model training as seen in Fig. 9.

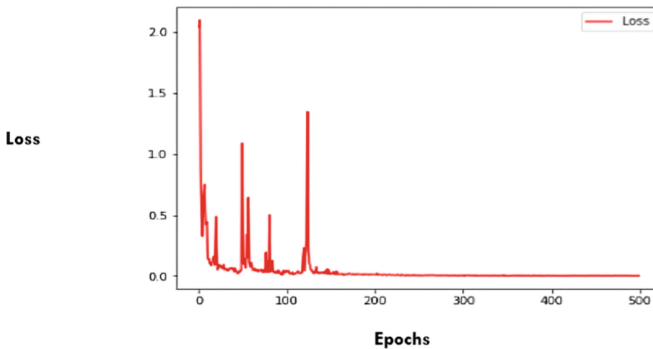


Fig. 9. Model Loss

The most important metric is accuracy, defined as the number of test triplets in which the anchor-positive distance is smaller than the anchor-negative distance. For each epoch, we calculate and print the accuracy, as well as the mean, standard deviation, minimum, and maximum values.

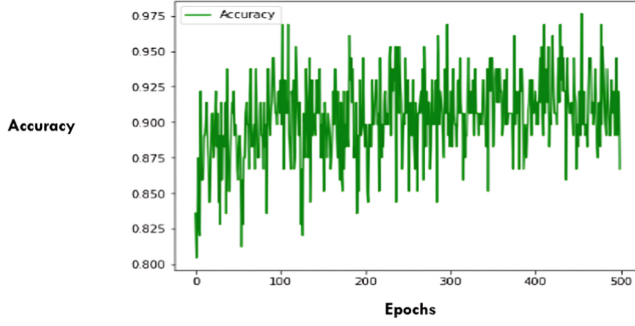


Fig. 10. Model Accuracy

This is done for 500 epochs, as shown in Fig. 10. The model successfully ran for 22797 s, which is approximately 6.33 h with max GPU. The accuracy peaks at 97.5%, making the Siamese Network a very good mechanism for face recognition from a single image. The addition of a ResNet50 base model also speeds up the training time significantly. The experiments were conducted on Google Colab with up to 12 GB of standard RAM and the Tesla T4 GPU.

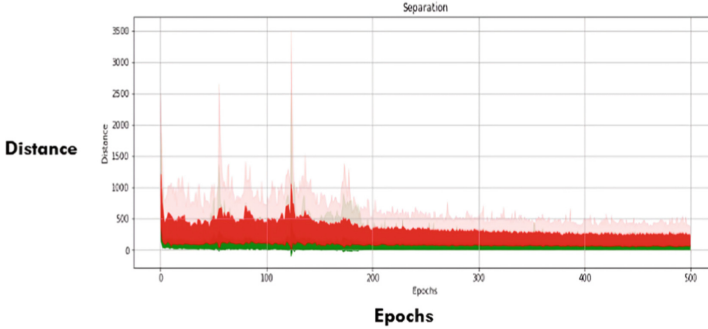


Fig. 11. Model Separation

The network provided good separation, which means that the distributions of the anchor-positive and anchor-negative distances will be non-overlapping (Fig. 11). This is an important result because it means that the network can distinguish between similar and dissimilar images.

After the training of the Siamese model is complete, we extract the weights into the encoder model and run it on all images to obtain their encoding vectors. We use these encodings to determine which people look alike. For this, we utilized the cosine similarity metric, defined as:

$$\text{Cosine Similarity} = \frac{u \cdot v}{\|u\| \cdot \|v\|} \quad (10)$$

where u and v are the encoding vectors of two images.

There were 143 people with more than 10 images. Of these, 115 (80.42%) were closest to themselves.

To evaluate the performance of our proposed method in the context of other approaches, we provide a comparative analysis in Table 1. This table highlights the datasets used, the methodologies employed, and the achieved accuracy for each method.

Table 1. Comparison With Related works

Paper	Dataset	Methodology	Accuracy
Robust Single-Sample Face Recognition [15]	LFW, AR	Sparse sub-dictionary learning with VGG-face net	97.1%
Deep Convolutional Neural Network for SSPP [3]	FERET, LFW, AR	Transfer learning with DCNN and expanded samples	98.3% (FERET), 96.2% (LFW)
Multi-Scale Features and Virtual Samples [4]	ORL, FERET	Virtual sample generation and SVM	95.2% (ORL), 94.8% (FERET)
Proposed Method	Celebrity	Siamese networks with triplet loss and preprocessing	97.5%

The proposed method achieves a competitive accuracy of 97.5%, demonstrating the effectiveness of the Siamese neural network approach in handling single-image face recognition tasks. The addition of preprocessing techniques such as segmentation, de-noising, and illumination correction further improved the performance by enhancing data quality.

5 Conclusion

In this paper, we proposed a mechanism to train face recognition systems using a single image per person by leveraging deep learning techniques. We identified image segmentation techniques that can be applied to a single image to generate multiple images, followed by preprocessing steps such as de-noising and illumination variation correction. Using Siamese neural networks, the proposed approach was evaluated on the Celebrity dataset and achieved a peak accuracy of 97.5%, demonstrating effective separation between positive and negative image pairs. This highlights the potential of Siamese Networks for face recognition tasks constrained by limited data. However, one key challenge was the high computational cost associated with training, which required significant memory and processing power, particularly for datasets with thousands of image pairs. Future work will aim to optimize the training process to reduce computational requirements and extend the model's applicability to more complex datasets and real-world scenarios.

References

1. Hafiz, F., Shafie, A.A., Mustafah, Y.M.: Face recognition from single sample per person by learning of generic discriminant vectors. *Procedia Eng.* **41**, 465–472 (2012)
2. Farfade, S.S., Saberian, M.J., Li, L.-J.: Multi-view face detection using deep convolutional neural networks. In: *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, pp. 643–650 (2015)
3. Zeng, J., Zhao, X., Gan, J., Mai, C., Zhai, Y., Wang, F.: Deep convolutional neural network used in single sample per person face recognition. *Comput. Intell. Neurosci.* **2018**(1), 3803627 (2018)
4. Li, F., Yuan, T., Zhang, Y., Liu, W.: Face recognition in single sample per person fusing multi-scale features extraction and virtual sample generation methods. *Front. Appl. Math. Stat.* **8**, 869830 (2022)
5. Tsai, T.-H., Chi, P.-T.: A single-stage face detection and face recognition deep neural network based on feature pyramid and triplet loss. *IET Image Proc.* **16**(8), 2148–2156 (2022)
6. Vargas, C., Zhang, Q., Izquierdo, E.: One shot logo recognition based on siamese neural networks. In: *Proceedings of the 2020 International Conference on Multimedia Retrieval*, pp. 321–325 (2020)
7. Xiang, W., He, R., Sun, Z., Tan, T.: A light CNN for deep face representation with noisy labels. *IEEE Trans. Inf. Forensics Secur.* **13**(11), 2884–2896 (2018)
8. Zheng, H.H., Zu, Y.X.: A normalized light CNN for face recognition. In: *Journal of Physics: Conference Series*, vol. 1087, p. 062015. IOP Publishing (2018)
9. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: a unified embedding for face recognition and clustering. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815–823 (2015)
10. Liu, L., Fieguth, P., Zhao, G., Pietikäinen, M., Dewen, H.: Extended local binary patterns for face recognition. *Inf. Sci.* **358**, 56–72 (2016)
11. Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**(1), 71–86 (1991)
12. Taigman, Y., Yang, M., Ranzato, M.A., Wolf, L.: Deepface: closing the gap to human-level performance in face verification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701–1708 (2014)
13. Wang, J., Plataniotis, K.N., Venetsanopoulos, A.N.: Selecting discriminant eigenfaces for face recognition. *Pattern Recognit. Lett.* **26**(10), 1470–1482 (2005)
14. Yu, Z., Zhang, C.: Image based static facial expression recognition with multiple deep network learning. In: *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pp. 435–442 (2015)
15. Cuculo, V., D’Amelio, A., Grossi, G., Lanzarotti, R., Lin, J.: Robust single-sample face recognition by sparsity-driven sub-dictionary learning using deep features. *Sensors* **19**(1), 146 (2019)



Development of Automated Banana Sorter Using Computer Vision Techniques

Semakaleng Seretloa , Pius Owolawi , and Chunling Tu 

Tshwane University of Technology, Pretoria, South Africa
smakalengs@gmail.com

Abstract. The main objective of this study is to develop and deploy an automated process of categorizing bananas according to their level of maturity or ripeness using computer vision techniques. This is done using advanced computer vision techniques with keen interest in exploring different types of transfer learning applied to pre-trained models. There is a lot of literature written about fruit-sorting systems using different technologies. This study focuses on the latest and most advanced machine learning concepts, and computer vision techniques which will produce a system able to perform efficiently in real-time. In this study banana dataset that is readily available online is used. The dataset is fed into a pre-trained model and results are studied. Different types of transfer learning are used to perform comparative analysis. Optimization techniques will be deployed to hyperparameters to foster high level accuracy performance. After model optimization, the model will be trained and then tested using the validation dataset. This paper provides an overview of existing literature, and results on how advanced computer vision techniques, coupled with transfer learning techniques, can be used to achieve high accuracy levels on pre-trained models.

Keywords: Computer Vision · Transfer Learning · Machine Learning · Pre-trained Model · Optimization

1 Introduction

The relatively recent spike in smart farming gave birth to the need to use more efficient methods of farming that will eliminate most of manual operations that were previously being executed by people. Machine Learning (ML) in general and Transfer Learning (TL) in particular, has generated a lot of discussions around the advancement of modernizing some of the traditional ways of farming and industrialization in general. Employing computer vision techniques to sort and grade products has been on the rise over time. This technique enables farmers to accurately distinguish products and give them a better insight into the products they are producing. This enables them to decide better when approaching the market. These new concepts can convert time-consuming and labor-intensive processes that require large amounts of data, computing power and resources, and often several iterations into seamless processes by just introducing pre-trained models which are optimized and fine-tuned using transfer learning techniques.

Transfer learning can improve the baseline performance of a model when increasing the knowledge of a separate case with one from a pre-trained model. This study offers an efficient way of sorting out bananas according to their level of maturity using computer vision techniques and transfer learning by applying fine-tuning techniques to a model. This paper experiments on a pre-trained ResNet101 model and presents highly accurate results after applying transfer learning to the model.

2 Literature

Automation of certain processes in agriculture has assisted farmers to control crop and livestock[1], and to handle production with better efficiency using minimum energy and low cost and to project the yield of production [2]. The scarcity of farming resources and the increased population reliance on agricultural production have necessitated researchers to have keen interest in developing automation systems to cater the growing need in the agriculture sector. Agricultural automation as indicated by [3, 4] has been executed by autonomous systems and machinery.

The concept of an automatic fruit sorter has long been of interest in agricultural crop production and preparing the produce for the market. Many studies were conducted to develop various fruit and vegetable classification systems using computer vision and image processing techniques. Researchers have been working on the design, development, and deployment of automatic systems for rapid and better yield estimation and quality grading for various crops. A widely adopted solution to automated fruit yield estimation is to use computer vision[5] to classify and detect defects in different fruits.

A study done to automatically sort the pili fruit was done [6] and achieved an overall average accuracy of between 95.5% and 98.0% using Linear Support Vector Machine(L-SVM) to classify three classes. The system could classify about 750 fruits in an hour. This yielded a much better result as compared to [7] which could only classify 507 fruits per hour with an efficiency of 96.97%. It would be interesting to see how a different model like ResNet with a step feeder conveyor system will perform.

According to [8], they found that using the YOLOv4 model to distinguish flowers from apples, and also giving the overall number of apples found in a tree produced an output of 70% to 88% accuracy.

In [9] authors use YOLOv3 to estimate the total production of citrus fruits and produced a precision of 0.96 with an average F1 score of 0.88. This suggests a good detection model. In [10], the researchers developed a Backpropagation Neural Network method to classify the maturity of strawberries, achieving high accuracy in their results.

Taking into consideration the population increase and rapid changes in climate, it is almost impossible to always be able to meet food requirements at any given point in time. Smart farming [11, 12] has come out to present itself as a creative way to solve the challenges currently experienced in food security and agriculture in general. Machine Learning (ML) is leading this innovative methods of trying to address the issues experienced in agricultural sustainability. The machines are taught and trained instead of the traditional way of instructing them programmatically on what to do.

In [13], researchers outline a wholistic survey on methodologies and challenges currently experienced in image processing looking at different approaches that researchers

take in addressing the subject of increasing productivity in an efficient way. Smart farming presents a good opportunity to explore what can be possible to maximize production without compromising the quality of the output produce.

Agricultural autonomous robots [14] have a role to play in food security of the future. Autonomous robots and their implementation in the industry has presented farmers with a strategic opportunity to fulfil farming operations promptly. Farming activities like planting and harvesting can now be executed efficiently with minimum labour and financial. This is due to the fact that the development of autonomous robots put main focus on imitating human labour. Different operations require different characteristics and have different specifications depending on the task to be executed and the expected output. There are still challenges that need to be investigated by researchers and more work is done to attack these challenges head-on. More and more robot are built specifically for farming industry to help the industry to thrive and produce optimal output ready for the consumption of the market.

An assessment study was done by [15] looking at post-harvest of bananas using a CNN-based ripeness classification model. What was interesting in the study is that the authors have developed a model that was tested using over 4000 images. The model achieved accuracy of 95%. Furthermore, they developed a webserver or application which is used as a dashboard to showcase the output of the model. The application can also be accessed from a smart mobile device, giving an element of convenience to the end-user.

Authors in [16] presented a study on classifying bananas depending on 6 stages of ripeness or maturity level. This is important because it gives commercial farmers insight into when to harvest and package depending of their clients. If the bananas still have to travel days to reach their destination, grading in this manner helps the packaging department to sort bananas according to their projected date of arrival, keeping in mind that the commercial distributor has to have time for repackaging.

In [17], a Recurrent Neural Network (RNN) architecture developed to work out instance segmentation and counting. The method's performance was presented using the dataset [18]. [19] shows how the Mask Regions with Convolutional Neural Networks (R-NN) method was tested in leaf segmentation. The authors of [20] approached the data augmentation with a Mask R-CNN network targeting the appearance characteristics of leaves.

A lot of literature has been written where researchers aim to implement object detection or segmentation networks to address the crop sorting problem. In this paper we consider all the literature available and put together a model that will produce high accuracy with efficient execution. The methodology approach, which is explained in the next section, details the approach used to achieve this study's objectives.

3 Methodology

In this study, suitable techniques and technologies are used to develop an accurate and robust system for transfer learning of image classification. The methods used include proper manipulation of the dataset for classification, data augmentation, data loaders, fine-tuning the models, model optimization and training as illustrated in Fig. 1.

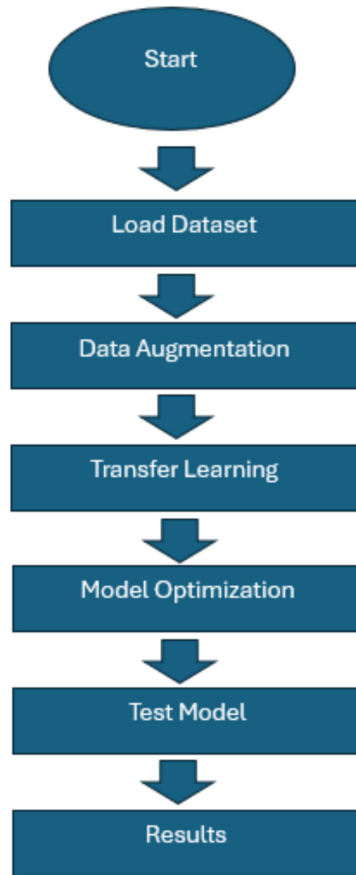


Fig. 1. Algorithm Flow Chart.

3.1 Dataset

The dataset used is split into training and validation datasets. Each dataset has three classifications; namely, unripe, fresh and rotten as illustrated in Fig. 2.

3.2 Dataloaders

Data loader is used to load data to make it easy to iterate through the dataset as needed. Once the data loaders are implemented for both training and validation datasets, a training model is defined using Cuda to access Google Processing Units (GPU).

3.3 Data Augmentation

When dealing with image datasets, some of the challenges often arise due to the inconsistencies in images such as, some being either too big or too small, and the limited number

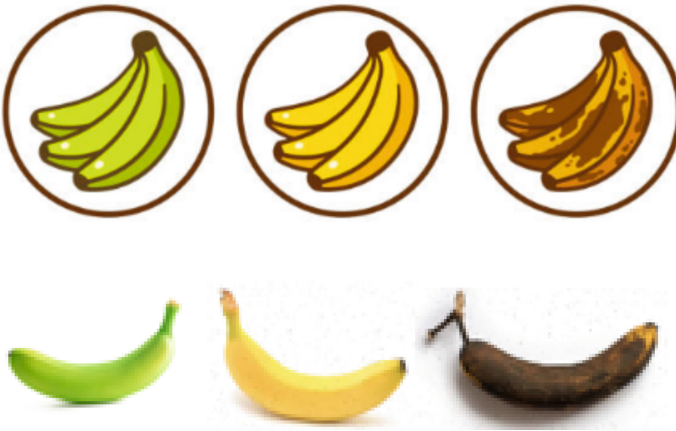


Fig. 2. Three classes of banana dataset.

of images in the training dataset which often results in overfitting. So, data augmentation is used to deal with these kind of issues.

Data augmentation strategies are utilized for all intents and purposes in most cutting-edge machine learning application such as localization, picture classification, recognition, normal dialect understanding, semantic division, and more. Augmented data moves forward the execution and brings about the profound learning models by creating modern and differing occurrences for preparing datasets.

In this study, data augmentation is utilized to convert the data or pictures to extend the capabilities of the model to perceive distinctive forms of an image. This process increments the breadth of data the model possesses.

Random cropping was used to transform the training dataset. The dataset was further transformed using horizontal axis flipping and then normalized. For the validation dataset, centre cropping was used to crop images to a specific width and height.

3.4 Transfer Learning

In Machine Learning, Transfer Learning is the re-use of a pre-trained model on a new problem. It is the enhancement of learning in a new but similar exercise through the exchange of information from a related work that has been taught utilizing pre-trained model. In this study, Transfer Learning was adopted because instead of reinventing a model from scratch, it is better to use the already learned knowledge of a pre-trained model on our dataset. This approach dramatically reduces computation time and resources required as compared to training a new model from scratch.

There are two methods of particular interest to perform transfer learning, which are discussed. The two methods are experimented with using the ResNet101 model. Table 1 shows the parameters used during testing.

Fine-Tuning the Pre-trained Model. Transfer learning was used by fine-tuning the pre-trained ResNet101 model. This was done by replacing the last fully connected layer

Table 1. Parameters used during experiments.

Parameters	Values
Epochs	20
Batch size	32
Learning rate	0.00001
Optimizer	Adam
Loss function	Cross entropy

for classifying and a typical CNN is depicted in Fig. 3 below. This method also applies fine-tuning on the pre-trained weights network by carrying on the backpropagation.

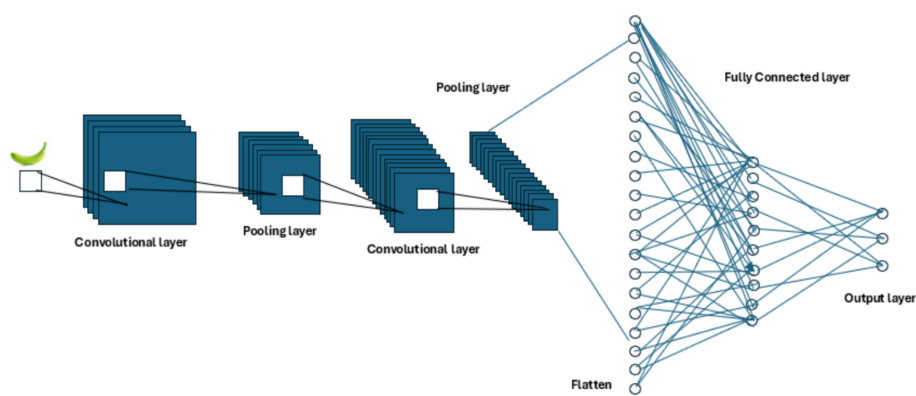


Fig. 3. A typical transfer learning model on a convolutional neural network.

Fixed Feature Extractor. Fixed feature extractor was used as an alternative to compare results. With this method, the final fully connected layer’s space is taken by a different layer with arbitrary weights. Training is only performed on this layer. All weights of the rest of the network are frozen or fixed. Only the weights of the last fully connected layer are not frozen. The remaining network is regarded as a fixed feature extractor for the new dataset after the last the last fully connected layer is removed.

3.5 Model Optimization

Developing a model that works effectively and makes accurate predictions in a certain set of situations is the main objective of machine learning. Machine learning optimization is required to accomplish it. The process of modifying hyperparameters to minimize the loss function using one of the optimization strategies is known as machine learning optimization.

For model optimization, three things have to be defined as explained below:

Optimization Algorithm with Learning Rate. Firstly, before the model can be trained, hyperparameters need to be established. The neural network's weights and learning rate are then changed using optimal function or algorithm. This increases the model's accuracy and lowers the overall loss. By reducing the inaccuracy while mapping inputs to outputs, it achieves this. The deep learning model's accuracy is significantly impacted by these optimizers or optimization methods. Depending on a particular problem, many optimization algorithms can be applied. This study employed the Adaptive Moment Estimation(Adam) technique. An adaptive learning rate approach called the Adam optimization algorithm was created to increase speed and hasten convergence in deep neural networks. It uses exponential moving average of the squared gradient (v_t), as in Root Mean Squared Propagation (RMSprop), in addition to the exponential moving average of the gradient (m_t). The following equations are used for parameter updates [21, 22]:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (1)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (2)$$

$$m_t^\wedge = \left(\frac{m_t}{1 - \beta_1^t} \right) v_t^\wedge = \left(\frac{v_t}{1 - \beta_2^t} \right) \quad (3)$$

$$\theta_{t+1} = \theta_t - \frac{n}{\sqrt{v_t^\wedge + \epsilon}} m_t^\wedge \quad (4)$$

where β_1 is the exponential decay rate. The default values for β_1 and β_2 are 0.9 and 0.999, respectively. $m_t^\wedge$ and $v_t^\wedge$ are correction biases for m_t and v_t respectively.

Set the Learning Rate Decay. Modern neural networks are trained using the learning rate decay technique. It begins with a high learning rate before repeatedly decaying it. It has been shown empirically to support both generalization and optimization. A high learning rate at first speeds up training or aids in the network's escape from erroneous local minima, whereas a decreasing learning rate aids in the network's convergence to a local minimum and prevents oscillation. A factor of 0.1 every seven epochs was used to establish the decay learning rate during testing in this paper.

Define Loss Function for the Model. A loss function needs to be established in order to train a machine learning model. What to seek for is a certain set of weights that will allow the neural network to produce accurate predictions, which will inevitably result in a lower loss function value. One way to assess how effectively the algorithm is modelling the dataset is to look at the loss function. It depends mathematically on the machine learning algorithm's parameters. A loss function is necessary to compute the model error during the optimization process used to train neural networks. The model's performance can be measured using the loss function. When training neural network models, the two primary loss function types to utilize are mean squared error and cross-entropy, and in this study, cross-entropy is used and the results are shown and discussed. The cross-entropy is calculated using Eq. 5 [23]:

$$H(t, p) = - \sum_{s \in S} t(s) \cdot \log(p(s)) \quad (5)$$

where $H()$ is the cross-entropy function, t may be the target distribution and p , the approximation of the target distribution

4 Results and Discussions

Figure 4 below displays the neural network’s error rate for training and validation datasets utilizing fine-tuning. An ideal solution is discovered in the 16th epoch, recording 0.03% validation loss and 100% accuracy. It is evident that training and validation errors reduce as the number of training iterations increases.

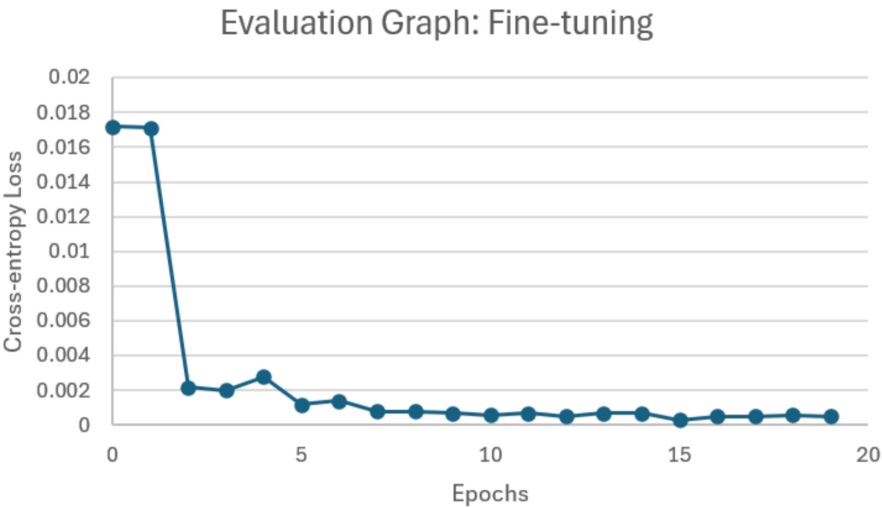


Fig. 4. Model performance after fine-tuning.

The quality of neural network predictions, as opposed to only the classification error, is measured via a cross-entropy plot. While the training error fell from 21.15% to 1.65% at iteration 14, the validation error decreased from 1.8% to 0.003% after 20 iterations. Fine-tuning was able to achieve an overall accuracy of 100% with no misclassification when testing. Table 2 below shows the performance evaluation of ResNet101 with set parameters and fine-tuning applied to the model.

Table 2. Fine-tuning performance evaluation.

Parameters	Values
Number of training images	4442
Number of validation images	997
Architecture	ResNet101
Accuracy	99.4

Transfer Learning by means of using a pre-trained ResNet101 model as a fixed feature extractor was experimented and the results are shown on Fig. 5 below. The experiment

yielded best validation overall accuracy of 82% at 15th epoch. The validation error over 20 epochs was reduced to 70%, whereas the training error was dropped to 79%. The results presented 5 out of 20 misclassification when tested over the validation dataset. This means the model only managed an accuracy of 75%.

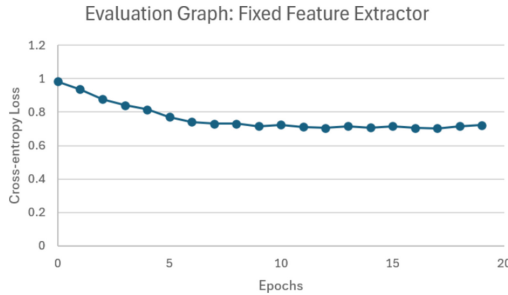


Fig. 5. Model performance after fixed feature extraction.

An experiment was conducted on the same dataset using Visual Geometry Group (VGG). It is a popular architecture known for its simplicity and effectiveness in image classification tasks. When tested on this experiment, it produced an overall accuracy of 69% over 20 epochs.

Another test was done using a custom-made convolutional neural network model and produced an overall accuracy of 95.5% over 30 iterations and 83% accuracy over 10 iterations. This suggests that the more iterations the model is trained on, the better the accuracy.

Transfer learning was used as a preferred method for this study because of its ability to use optimization algorithms for improved efficiency and accuracy. It can be deduced that the best-performing transfer learning technique is fine-tuning the pre-trained model.

5 Conclusion

This paper introduced the concept of using Transfer Learning as a tool to develop a model that will be able to sort out and categorize bananas according to their level of maturity or ripeness. The experiment used a dataset which was divided into training and validation datasets. The pre-trained model was trained using the training dataset, and the model was subsequently tested using the validation dataset.

From the experiment when using Fine-tuning of a pre-trained model, results of a high accuracy level of between 99.4% to 100% are achieved.

A brief comparison was conducted between the effects of using either a Fixed Feature Extractor or Fine-tuning pre-trained model and the results show that the latter achieve more desirable results.

This paper only focuses on designing the algorithm to achieve more accurate results. A full architecture system has not yet been developed. Future work may include building a real-time system on a conveyor belt and using Internet of Things (IoT) concepts such as Digital Twin Framework.

References

1. Micle, D.E., et al.: Research on innovative business plan. Smart cattle farming using artificial intelligent robotic process automation. *Agriculture* **11**(5), 430 (2021)
2. Bazame, H.C., Molin, J.P., Althoff, D., Martello, M.: Detection, classification, and mapping of coffee fruits during harvest with computer vision. *Comput. Electron. Agric.* **183**, 106066 (2021)
3. Defourny, P., et al.: Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. *Remote Sens. Environ.* **221**, 551–568 (2019)
4. Rose, D.C., Lyon J., de Boon, A., Hanheide, M., Pearson, S.: Responsible development of autonomous robotics in agriculture. *Nat. Food* **2**(5), 306–309 (2021)
5. Kumar, S.D., Esakkirajan, S., Bama, S., Keerthiveena, V.: A microcontroller based machine vision approach for tomato grading and sorting using SVM classifier. *Microprocess. Microsyst.* **76**, 103090 (2020)
6. Novelero, J.M., Cruz, J.C.D.: Automated pili fruit sorting device using L-SVM image classification model
7. Radores, L.C., Taylan, V.T., Sicat, E.V., Lavarias, J.A.: Development and evaluation of Pummelo (*Citrus grandis* L.) fruit sorter
8. Lee, J., Gadsden, S.A., Biglarbegian, M., Cline, J.A.: Smart agriculture: a fruit flower cluster detection strategy in apple orchards using machine vision and learning. *Appl. Sci.* **12**(22), 11420 (2022)
9. Vijayakumar, V., Ampatzidis, Y., Costa, L.: Tree-level Citrus yield prediction utilizing ground and aerial machine vision and machine learning. *Smart Agric. Technol.* **3**, 100077 (2023)
10. Wang, X., Tu, C., Owolawi, P.A.: Strawberry maturity classification based on BP neural network. In: *Smart Trends in Computing and Communications: Proceedings of SmartCom 2021*, pp. 311–321. Springer (2022)
11. Khanna, A., Jain, S., Maheshwari, P.: Precision agriculture for medicinal plants: a conjunction of technologies. In: *2022 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, pp. 300–304. IEEE (2022)
12. Sharma, A., Jain, A., Gupta, P., Chowdary, V.: Machine learning applications for precision agriculture: a comprehensive review. *IEEE Access* **9**, 4843–4873 (2020)
13. Mayuri, K., Priya, C.: Role of image processing and machine learning techniques in disease recognition, diagnosis and yield prediction of crops: a review. *Int. J. Adv. Res. Comput. Sci.* **9**(2) (2018)
14. Mahmud, M.S.A., Abidin, M.S.Z., Emmanuel, A.A., Hasan, H.S.: Robotics and automation in agriculture: present and future applications. *Appl. Model. Simul.* **4**, 130–140 (2020)
15. Arunima, P., Gopinath, P.P., Lekshmi, P.G., Esakkimuthu, M.: Digital assessment of post-harvest Nendran banana for faster grading: CNN-based ripeness classification model. *Postharvest Biol. Technol.* **214**, 112972 (2024)
16. Kosasih, R., Sudaryanto, S., Fahrurrozi, A.: Classification of six banana ripeness levels based on statistical features on machine learning approach. *Int. J. Adv. Appl. Sci.* **12**(4), 317 (2023)
17. Ren, M., Zemel, R.S.: End-to-end instance segmentation with recurrent attention. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6656–6664 (2017)
18. Minervini, M., Fischbach, A., Scharr, H., Tsaftaris, S.A.: Finely-grained annotated datasets for image-based plant phenotyping. *Pattern Recogn. Let.* **81**, 80–89 (2016)
19. Xu, L., Li, Y., Sun, Y., Song, L., Jin, S.: Leaf instance segmentation and counting based on deep object detection and segmentation networks. In: *2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS)*, pp. 180–185. IEEE (2018)

20. Kuznichov, D., Zvirin, A., Honen, Y., Kimmel, R.: Data augmentation for leaf segmentation and counting tasks in rosette plants. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (2019)
21. Sun, S., Cao, Z., Zhu, H., Zhao, J.: A survey of optimization methods from a machine learning perspective. *IEEE Trans. Cybern.* **50**(8), 3668–3681 (2019)
22. Ruder, S.: An overview of gradient descent optimization algorithms. arXiv preprint [arXiv: 1609.04747](https://arxiv.org/abs/1609.04747). (2016)
23. Zhou, Z., Huang, H., Fang, B.: Application of weighted cross-entropy loss function in intrusion detection. *J. Comput. Commun.* **9**(11), 1–21 (2021)



Evaluating the Effectiveness of Current Countermeasures Against Man-In-The-Middle Attack on Domain Name System

Phandavhudzi Rotshidzwa^(✉), Moila Ramahlapane Lerato, and Velempini Mthulisi

Department of Computer Science, University of Limpopo, Polokwane, South Africa
Rotshi06@gmail.com, {lerato.moila,mthulisi.velempini}@ul.ac.za

Abstract. In an advancing digital environment, the Domain Name System (DNS) serves as a significant component facilitating Internet functionality. However, it is susceptible to Man-in-the-Middle (MitM) attacks which compromise security and privacy. We investigate the MitM attacks and propose a security measure in addition to authentication to mitigate the effects of attacks. The literature review highlights the need for DNS monitoring and detection techniques to address challenges such as encrypted traffic and false positives. This study proposes novel approaches to enhance DNS security against MitM attacks. The Simulink module in MATLAB was employed to simulate and model DNS systems and network configurations, enabling the design of simulation models which replicate various attack scenarios and testing countermeasures' effectiveness. The Detection Accuracy, Packet Drop Ratio, DNS Response time, Packet Delivery Ratio, Throughput, and Average Delay were considered for the evaluation of the proposed scheme. Results show that the Location-Based model demonstrates superior performance. The scheme performed well in detection accuracy, packet delivery ratio, average response time, and throughput compared to DNSSEC and CGUARD. The scheme is effective in mitigating MitM attacks. The Location-Based model is recommended as an effective defence mechanism against MitM attacks. Ongoing modifications and proactive measures are recommended to safeguard Internet security.

Keywords: domain name system · man-in-the-middle attacks · encryption · authentication

1 Introduction

The Domain Name System (DNS) is a fundamental and key technology in the Internet's infrastructure, facilitating the seamless translation of user-friendly domain names into Internet Protocol (IP) addresses. This translation process is crucial for accessing online resources, including processing of financial transactions, e-commerce, and sensitive personal and professional data [1]. However, the reliance on DNS also results in security risks, as it is susceptible to various cyber threats, particularly Man-in-the-Middle (MitM) attacks. In such attacks, unauthorized third parties intercept and modify

data during transmission or redirect users to malicious websites without their knowledge [2]. This poses significant risks to the confidentiality, integrity, and availability of online interactions, as illustrated in Fig. 1.

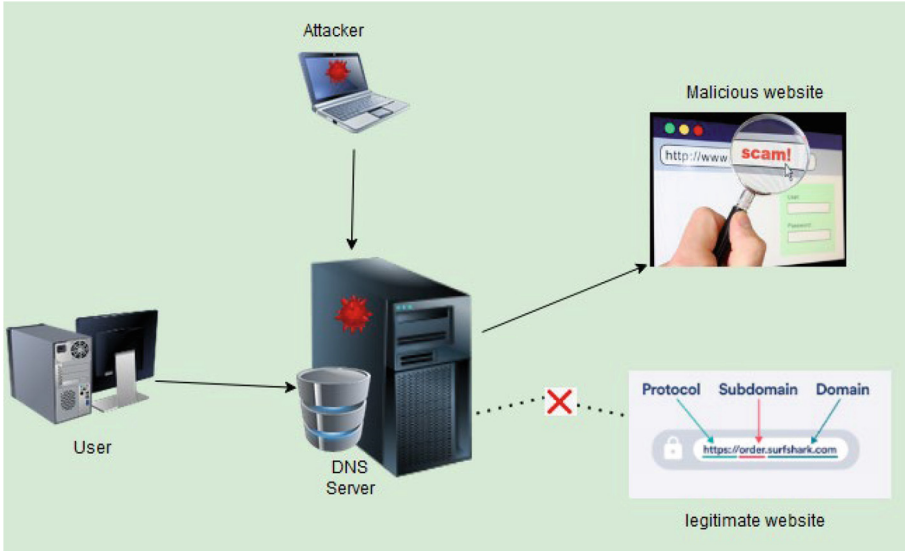


Fig. 1. MitM attack redirects users to a fake website.

With the rise of cyber threats, understanding the severity of MitM attacks on the DNS is significant. As the backbone of the Internet, DNS facilitates the translation of domain names into IP addresses, enabling seamless access to websites and online resources. However, DNS has become increasingly vulnerable to security attacks such as frequent MitM attacks that compromise the confidentiality, integrity and availability of web-based services [3]. Through these attacks, malicious actors intercept and manipulate DNS queries, often redirecting users to fraudulent sites that mimic legitimate ones. This can lead to significant breaches of sensitive information, particularly in finance and e-commerce, where users unknowingly submit personal data to attackers. Thus, there is a need for innovative strategies that not only authenticate DNS users but also minimize the impact of MitM attacks [4].

From 2020 to 2024, South Africa has faced significant cyber threats, including DNS-based MitM attacks. In 2021, Interpol reported that South Africa had the third-highest number of cybercrime victims globally, with substantial financial losses. The period from 2022 to 2024 saw a 37% increase in cyber-attacks, often involving sophisticated DNS manipulation techniques. Reports in 2024 highlighted DNS spoofing and cache poisoning as major threats, where attackers insert corrupted DNS data to redirect users to malicious sites. Additionally, Investec's cybersecurity insights for 2024 emphasized the growing sophistication of cyber threats, including advanced phishing and malware attacks that can facilitate MitM scenarios.

Several techniques have been proposed to address DNS security vulnerabilities, with DNS Security Extensions (DNSSEC) being a widely used mechanism to mitigate MitM attacks on DNS [3]. By digitally signing DNS data, DNSSEC enhances the security of DNS ensuring its authenticity and integrity. However, while DNSSEC and similar authentication-based measures have significantly improved DNS security, challenges remain. Attackers continue to exploit weaknesses that authentication alone cannot fully mitigate. Furthermore, interception and redirection tactics that reroute users to malicious sites may not be detected. To address these challenges, there is a growing need for multi-facet strategies that go beyond authentication to actively detect and minimize the impact of MitM attacks on DNS [5].

This research study investigates and proposes innovative security approaches that not only authenticate DNS data but also enhance detection and mitigation against interception. By building on existing frameworks and addressing the limitations of current solutions, the study aims to contribute to a more robust DNS security model that strengthens user data protection and ensures secure online transactions [6]. In Sect. 2, the study focuses on the related works that will explore limitations and methods utilized to date, Sect. 3 focuses on the methods, and tools used to carry out the experiments, Sect. 4 discusses the simulation results and Sect. 5 concludes the study.

2 Related Work

Literature has shown that techniques exist which can be used to launch a MitM attack on the DNS, such as DNS poisoning, which reroutes traffic by inserting a wrong hostname in the victim's DNS cache. On the other hand, there exist tools such as the Ettercap and SSL-strip, which are included in Kali Linux, one of the strongest security packages available for ethical hackers which was created by Mati Aharoni and Devon Kearns [7]. The authors only demonstrated the attack and concluded that a user must be alert when using a seemingly secure website such as https:// or the one which displays a locked padlock. Attackers can use SSL-strip to create fake SSL which use https:// or locked padlock [8].

DNSSEC is a suite of protocols developed by the IETF community in response to threats to the DNS infrastructure including MitM [9]. DNSSEC assures the authenticity of the origin of the response and provides integrity for the data received from DNS servers. The authors observed that while a MitM attack targeting a DNSSEC client remains active, it cannot compromise a recursive server. This means that a falsified DNSSEC response will fail to be validated by the recursive server, but the client may still not receive the accurate DNSSEC response [10].

In [11], vulnerabilities in DNS infrastructure which can be exploited by MitM attacks such as cache poisoning attacks which is also known as DNS spoofing are explored. DNS spoofing is an informational-level attack, which compromises the DNS server. DNS Hijacking is a type of attack where an attacker gains unauthorised access to a domain's DNS configuration and modifies the domain's IP address and redirects users to a malicious website. The work in [11], summarises all possible attacks on the DNS infrastructure and those attacks which exploit the DNS infrastructure. Furthermore, countermeasure techniques such as the DNSSEC, TSIG and DNS firewalls are discussed.

However, according to the authors, the techniques cannot protect DNS from all attacks. There is a need for secure DNS techniques designed to counter all DNS attacks.

The challenges associated with DNS monitoring and detection techniques present significant obstacles in the endeavour to mitigate MitM attacks in DNS. The increasing use of encryption technologies like DNS-over-HTTPS (DoH) and DNS-over-TLS (DoT) has improved the security of DNS traffic which protects data in transit from eavesdropping and tampering. Unfortunately, with DNS encryption, monitoring becomes a challenging task. One cannot inspect and detect potential threats within the encrypted data. Consequently, innovative detection methods may be designed specifically for encrypted traffic or focus on the detection of suspicious behavioural patterns associated with DNS-based attacks [12].

Additionally, the false positives in DNS monitoring and detection create a significant operational challenge. False positives occur when legitimate DNS traffic is mistakenly categorized as malicious, flooding security teams with an overwhelming volume of alerts. Addressing this challenge is imperative to ensure that security resources are effectively allocated to real threats rather than wasted on false alarms. To counter zero-day attacks, which do not have known threat signatures, researchers may employ heuristic and behaviour-based approaches which can detect anomalies and abnormalities from typical DNS traffic patterns.

Furthermore, given the dynamic nature of DNS infrastructure, characterized by frequent changes in IP addresses and domain names, the development of adaptive monitoring systems becomes fundamental. These techniques must adapt quickly to evolving DNS infrastructure and provide high detection accuracy as the DNS technology advances. In addressing these challenges, the goal is to enhance the effectiveness of DNS monitoring and detection techniques thereby improving the security of online transactions and data transmission [13].

In study in [14] presents a survey of DNS vulnerabilities and Attacks. The survey is categorized into three categories of DNS vulnerabilities which are conceptual, structural and communication based. The study further discusses DNSSEC vulnerabilities and concludes that DNSSEC has enhanced security for authentication and integrity by incorporating digital signatures using public and private key infrastructure to existing DNS systems to counter known DNS vulnerabilities. However, DNSSEC is still vulnerable to various attacks which compromise those vulnerabilities and limitations such as MitM attacks. In [15], authors noted that even though DNSSEC still has vulnerabilities and limitations, it is more secure than DNS. DNSSEC still has limitations because DNS security extensions provide original authentication and integrity protection for DNS data, as well as a means of public key distribution [16].

These extensions do not provide confidentiality as observed in [15] and [16].

Confidentiality is very important on the Internet because without confidentiality, DNS data is compromised. DNSSEC is a suite of protocols developed by the Internet Engineering Task Force (IETF) community in response to DNS threats which includes MitM [17]. DNSSEC assures the authenticity of the origin of the response and provides integrity of received DNS data. As observed in [10] though DNSSEC has limitations, falsified DNSSEC response cannot be validated by the recursive server. Unfortunately, the client may still not receive the correct DNSSEC response [18]. The works in [10] and

[18] provide a list of factors to be considered in improving DNS security and evaluating the effectiveness of existing MitM attack countermeasures in DNS.

3 Methodology

For Simulation and Modelling, MATLAB's Simulink module was used to simulate and model the behaviour of DNS systems and to configure the network. We designed simulation models to replicate various attack scenarios and evaluate the countermeasures under controlled conditions. Simulink's visual modelling environment can model a DNS system, the attacks and the countermeasures. Results can be represented graphically to aid in the analysis of the interactions of DNS systems and attacks effectively. The evaluation metrics which were considered include Detection Accuracy, Packet Drop Ratio (PDR), DNS Response time, Packet Delivery Ratio, Throughput, and Average Delay. The evaluation metrics are important in providing and evaluating the effectiveness of countermeasures in countering the MitM attacks. We considered a network grid which is 500 by 500 m.

The performance of the following mechanisms DNSSEC, CGUARD, and Location Based were compared. Email servers use DNS to route messages after resolution, which means they are vulnerable to security issues in the DNS infrastructure. In September 2014 researchers at CMU found that emails supposed to be sent through Yahoo, Hotmail, and Gmail servers were routed through rogue mail servers. Attackers exploited a decades-old vulnerability in the DNS. The Solution was to utilize the DNSSEC, which adds a layer of trust on top of DNS by providing authentication [19]. While on the other hand, CGUARD is a promising DNS security tool designed to enhance the detection and prevention of DNS-based attacks. It combines anomaly detection with real-time monitoring, targeting MitM and other threats. It was chosen for this study due to its advanced detection capabilities and its effectiveness in environments requiring robust, adaptive security against DNS manipulation [20].

The Location-based model uses geolocation and IP-based verification to prevent unauthorized DNS responses by ensuring that DNS queries and responses originate from expected locations. This model is particularly effective in identifying abnormal patterns associated with MitM attacks, making it a valuable addition to this study as it combines both simplicity and specificity in its approach to threat prevention. The simulation framework consisted of a standardized duration of 10 s for all scenarios. To ensure consistency across the simulations, a Query Arrival Rate was set to 2 queries per second. Our simulation environment, common to all scenarios, consisted of two DNS servers, two clients, and one MitM attacker. This configuration enables one to conduct a thorough and standardized evaluation of the selected DNS security mechanisms. Our Scenario consists of two elements, the clients and two DNS servers. It is within this environment, that we conducted a series of simulations to evaluate the impact of various security mechanisms in the DNS.

Our simulation began with a baseline simulation in which we ran the DNS without any protective mechanism. This enabled us to establish a reference point to determine how security mechanisms subsequently impact the performance of the DNS. The baseline is also essential in evaluating whether the mechanisms enhance or degrade DNS functionality and security. The security mechanisms were implemented in the DNS servers one

at a time. DNSSEC was the first to be implemented followed by the CGUARD and the Location-Based model respectively.

4 Experimental Results

In this section, we focus on the experimental simulation to evaluate the performance of the security mechanism designed to secure DNS and mitigate against MitM attacks. Figure 2 shows the comparative detection accuracy results of DNS, DNSSEC, CGUARD and LOCATION-BASED models. In Fig. 2, we can observe the performance of the security mechanisms regarding DNS detection accuracy. Figure 2 demonstrates whether these mechanisms either enhance or degrade the overall security of the DNS infrastructure. Considering DNS without any protective mechanism, it shows the baseline - the performance of DNS without protective mechanisms. The accuracy level reflects that traditional DNS lacks advanced mechanisms to detect and counter MitM attacks effectively. It implies that standard DNS alone is insufficient for environments where high security is required.

The standard DNS requires minimal computational resources because it does not perform complex security checks. However, this simplicity also limits its security capabilities though it performs better than DNSSEC and CGUARD. While the LOCATION-BASED model is superior in improving the DNS security. It improved detection accuracy by 52,5%, which depicts the capability of LOCATION-BASED in improving DNS security against MitM attacks. The model optimises effectiveness with relatively moderate computational demands, suggesting it could be a practical solution for environments that require robust security without high computational overhead.

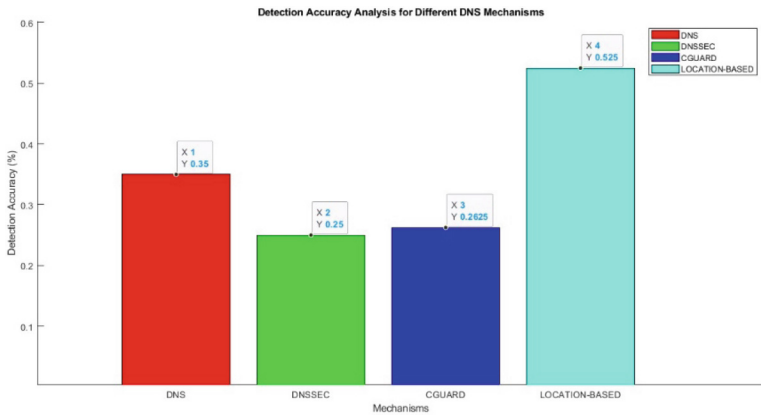


Fig. 2. Comparative Analysis of Detection Accuracy

Figure 3 presents the delay results of the DNS, DNSSEC, CGUARD and LOCATION-BASED models. Incorporating the basic DNS configuration without any security features allows one to evaluate how security measures impact the latency in DNS procedures. The results show the extent to which these mechanisms either enhance

or affect the data transmission speed within the DNS infrastructure. Without the protective mechanism, DNS exhibits an average of 0 ms because DNS does not perform any additional security checks or cryptographic processes, which keeps query processing extremely fast thus DNS's main purpose is to translate domain names to IP addresses without authentication or verification, resulting in almost zero delay.

When the DNSSEC is implemented, there is an increase of 0.079 ms. DNSSEC involves signing and verifying DNS responses to authenticate their integrity. This cryptographic process requires additional computational resources and time, leading to increased latency. While for CGUARD the delay is 0.064 because CGUARD uses real-time monitoring and anomaly detection algorithms, which require analyzing traffic patterns. This process introduces delays, as each query may be compared to the previous patterns to detect anomalies, particularly in networks with high traffic volumes. CGAUARD utilizes more advanced detection techniques to detect potential MitM or other DNS-based attacks. However, it requires extra processing which affects the response times.

While the LOCATION-BASED maintains a good performance with an average delay of 0 ms. This shows the good performance of the LOCATION-BASED model in maintaining the efficiency of DNS operations as shown in Fig. 3.

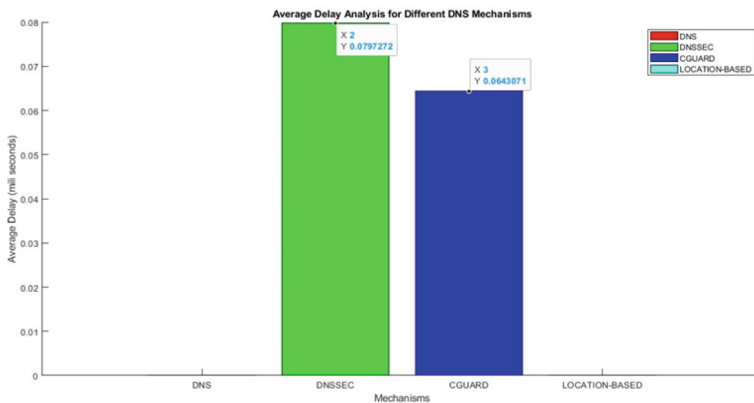


Fig. 3. Comparative analysis of average delay results.

Figure 4 presents a comprehensive comparative analysis of PDR results of DNS, DNSSEC, CGUARD and the LOCATION-BASED model. The results show the impact of these security measures on DNS. The baseline DNS, without security mechanisms, outperformed DNSSEC by 35% which achieved a 25% PDR. This lack of processing overhead results in higher PDR, though it comes at the expense of security while DNSSEC improves security, the cryptographic operations reduce PDR by increasing response times. The CGUARD minimizes PDR by 26.25%, it uses real-time monitoring and anomaly detection while enhancing security. It also consumes network resources

and adds complexity to DNS traffic handling, resulting in lower PDR. The LOCATION-BASED demonstrated significant improvements with 52% PDR, highlighting its effectiveness in enhancing DNS reliability. This analysis is important for network administrators and security experts who are attempting to improve DNS security while achieving high PDR and reliability as shown in Fig. 4.

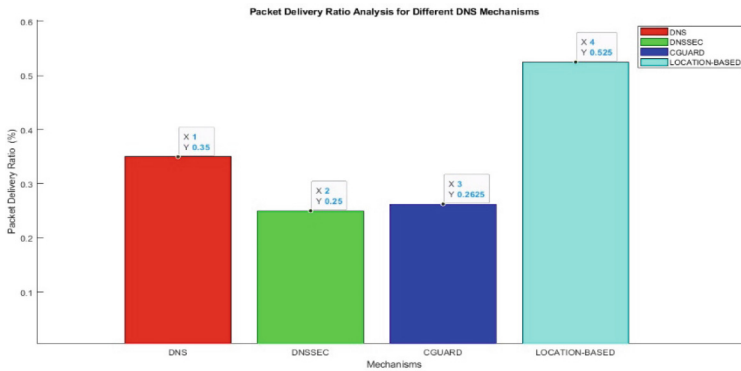


Fig. 4. Comparative analysis of packet delivery ratio results.

A comparative analysis of the Packet Drop Rate of DNS, DNSSEC, CGUARD, and LOCATION-BASED models shows the reliability of the models. In the absence of protective mechanisms, DNS exhibited a 50% Packet Drop Rate, while the Introduction of DNSSEC resulted in a lower 36.66%, CGUARD 41.25% and LOCATION-BASED with 30%. The mechanism primarily focuses on speed and ease of packet transmission without any additional processing. The results emphasize the diverse effects of security mechanisms on packet transmission reliability within the DNS, providing valuable insights into trade-offs and challenges in securing the DNS environment and prompting further investigation and enhancement of security measures as depicted in Fig. 5.

Figure 6 depicts an analysis of false positive rates of DNS, DNSSEC, CGUARD, and the LOCATION-BASED models. Figure 6 presents insights into the performance security mechanisms in terms of false positives in comparison to the Packet Drop Rate in Fig. 5. Without security mechanisms, DNS incurred 50% Packet Drop Rate and 61.53% false positives. DNSSEC minimized the drop rate to 36% with 71.45% false positives while CGUARD incurred a 41.25% drop rate with 71% false positives. The LOCATION-BASED MODEL incurred a 30% drop rate with 52.95% false positives, which highlights its reliability. This analysis underscores each mechanism's ability to minimize packet drop rate and, in this regard, the LOCATION-BASED model was superior.

Figure 6 presents a comparative analysis of False Positive Rates of DNS, DNSSEC, CGUARD, and the LOCATION-BASED MODEL. The analysis highlights the reliability of each security mechanism in mitigating packet drop rate and preserving DNS integrity. Without any security mechanism, 61.53% of dropped packets were observed to be false positives, which shows that the system is not reliable. Despite a reduced overall packet drop rate with DNSSEC and CGUARD, a significant percentage of these dropped packets were still false positives. Notably, the LOCATION-BASED MODEL demonstrated good

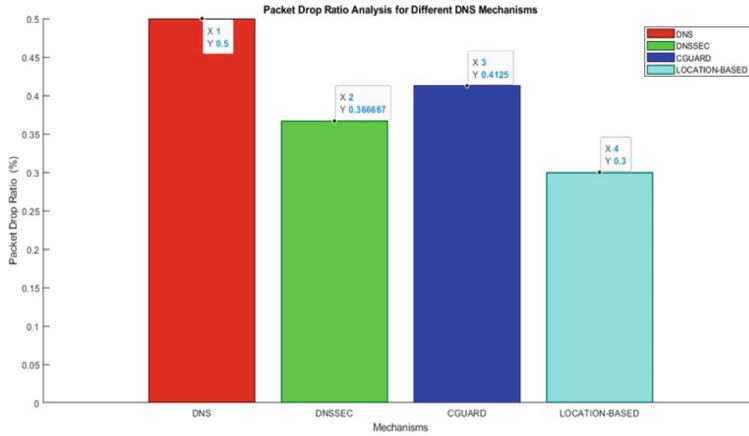


Fig. 5. Comparative analysis of packet drops ratio results.

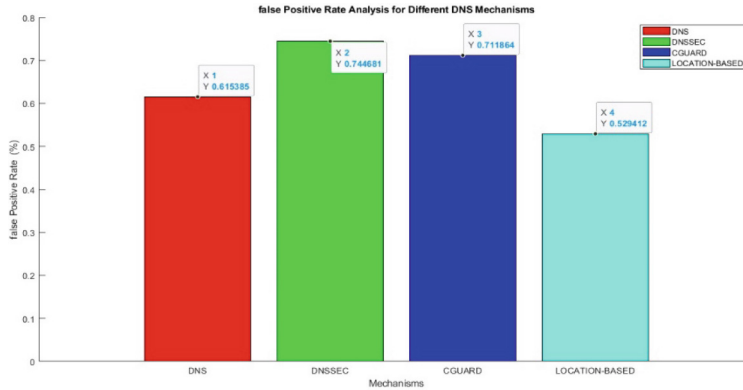


Fig. 6. Comparative analysis of false positive rate results.

performance, achieving a 30% packet drop rate with 52.95% being false positives. This underscores the LOCATION-BASED model as a dependable solution that trades off security and operational efficiency.

In Fig. 7, a comprehensive comparative analysis of Throughput for DNS, DNSSEC, CGUARD, and the LOCATION-BASED MODEL is conducted, providing crucial insights into the efficiency of data transmission within the DNS under different security mechanisms. These results shed light on how each security measure directly influences Throughput, revealing their effects on the successful transmission of data within the DNS infrastructure. Without protective mechanisms, DNS demonstrated a commendable 40% Throughput, but the introduction of DNSSEC and CGUARD resulted in a significant reduction to 20%. In contrast, the LOCATION-BASED MODEL exhibited superior performance with an impressive 50% Throughput, indicating its potential to enhance data transmission efficiency while ensuring security. The detailed analysis in Fig. 7 highlights

the diverse impacts of these security mechanisms on successful data transmission rates in the DNS, offering valuable insights for decision-makers and network administrators to make informed choices that optimize DNS operations while considering the trade-offs between security and performance.

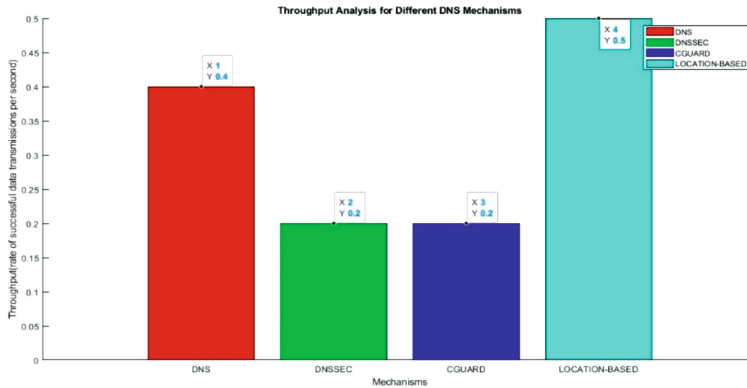


Fig. 7. Comparative analysis of throughput results.

Figure 8 presents a comparative analysis of the Average Response Time of DNS, DNSSEC, CGUARD, and the LOCATION-BASED model which captures the responsiveness of the DNS under different security measures. The results highlight the influence of each security mechanism on Average Response Time which shows how the security mechanisms impact the speed and efficiency of DNS resolution. Without security mechanisms, DNS exhibited an Average Response Time of 5.40/s. The introduction of DNSSEC improved response time to 4.64/s, which shows its positive impact on DNS efficiency. In contrast, CGUARD resulted in a longer Average Response Time of 7.65/s, while the LOCATION-BASED model had a response time of 8.09/s which is slightly higher than CGUARD. Figure 8’s analysis of Average Response Time underscores the impact of these security mechanisms on DNS resolution speed. It also provides insights into how DNSSEC can optimize response times and facilitate a comparative analysis that highlights the significance of trading off security and DNS performance.

Table 1 provides a comparative analysis of DNSSEC, CGUARD, and the Location-Based models based on the results in Figs. 2, 3, 4, 5, 6, 7 and 8. The Location-Based model was superior in six of the seven scenarios due to its unique geographical-based authentication method. This method ensures that only the DNS server in the location of the query responds, resulting in higher detection accuracy and occasional false positives during location changes. The model’s approach also contributes to reduced average response times. Notably, the Location-Based model is effective against MitM attacks. In contrast, DNSSEC, focusing on data encryption and authentication depicted an average response time though it was outperformed by the Location-Based model given its efficacy. CGUARD, a cache-based solution, underperformed across all metrics due to its vulnerability to attacks not present in the cache, leading to reduced overall performance.

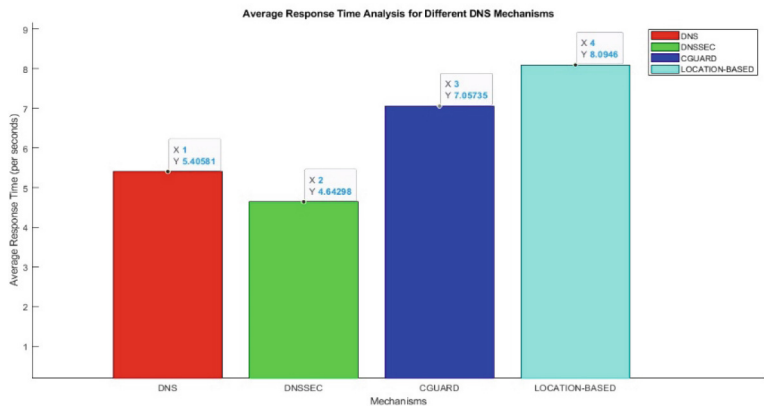


Fig. 8. Comparative analysis of average response time results.

Table 1. Analysis of the Performance of the Security Mechanism in all the Scenarios.

	AVG Delay	Detection Accuracy	PDR	Loss Packets	Avg Re-sponse time	False Positive	Through put
MECHANISMS	DNSSEC				✓		
	CGUARD						
	LOCATION BASED MODEL	✓	✓	✓	✓		✓

5 Conclusion

In conclusion, this study provided insights into the impact of MitM attacks on the performance of DNS, highlighting the significance of Internet security and user privacy. Through the experimental simulations in which many DNS security mechanisms such as the DNSSEC, CGUARD, and the Location-Based model, effective security strategies designed to mitigate security attacks were highlighted. The findings underscore the superior efficacy of the Location-Based model in improving DNS security, suggesting that this model could be adopted as a robust defense against MitM attacks.

DNSSEC has certain limitations that affect its effectiveness. For example, DNSSEC’s performance overhead due to cryptographic operations can lead to increased latency and higher data traffic. DNSSEC’s complexity and key management requirements make it challenging to maintain. It is more susceptible to misconfigurations, which can disrupt legitimate traffic.

Optimizations may adopt more lightweight cryptographic algorithms, like elliptic curve cryptography, which reduces response times and processing loads. Improvements in automated key management could simplify DNSSEC's deployment and reduce its complexity while reducing instances of misconfiguration. The deployment of models that allow hybrid DNSSEC setup may also promote wider adoption of DNSSEC by many different organizations. Future research may prioritize the enhancement and optimization of the Location-Based model, given its demonstrated benefits in detection accuracy and packet delivery.

The study was limited to a simplified experimental setup consisting of two DNS servers, two client machines, and a single MitM attacker, with each simulation limited to a fixed duration of 10 simulation seconds. While this setup provided initial insights into the effectiveness of the evaluated security mechanisms, it does not capture the full complexity of real-world DNS environments, which often involve numerous servers, diverse client configurations, and varying attack vectors.

Furthermore, investigations are required into innovative approaches that could help in addressing emerging threats in the dynamic landscape of Internet security. Proactive measures are essential to counter evolving cyber threats, ensuring the resilience and integrity of the DNS infrastructure. On the other hand, advancing DNS security mechanisms, like DNSSEC and Location-Based models, is vital to maintaining a secure and reliable Internet ecosystem.

Availability Statement. The security of the Domain Name System (DNS) infrastructure is critical in preventing unauthorized interception and manipulation of DNS traffic. In our research, we assess the efficacy of existing countermeasures against man-in-the-middle attacks targeting DNS. These countermeasures include DNSSEC, CGUARD, and location-based models. By analyzing their impact on security, we aim to identify whether they enhance or degrade the overall resilience of the DNS ecosystem. For further inquiries or collaboration, please feel free to reach out to me at moilaramahlapane@gmail.com.

References

1. Zhang, Y.L., Xia, G.S.: The SSL MITM attack with DNS spoofing. *Appl. Mech. Mater.* **385–386**, 1647–1650 (2013). <https://doi.org/10.4028/www.scientific.net/AMM.385-386.1647>
2. Wang, Y., Yun, X., Xiong, G., Li, Z., Yao, Y.: Availability analysis of DNSSEC resolution and validation service. In: 7th International Conference on Communications and Networking in China, pp. 35–42 (2012). <https://doi.org/10.1109/ChinaCom.2012.6417444>
3. Tzur-David, S., Lashchiver, K., Dolev, D., Anker, T.: Delay Fast Packets (DFP): prevention of DNS cache poisoning. In: *Security and Privacy in Communication Networks*, pp. 303–318. Springer (2012). https://doi.org/10.1007/978-3-642-31909-9_17
4. Mummadi, A., Yadav, B.M.K., Sadhwika, R., Shitharth, S.: An appraisal of cyber-attacks and countermeasures using machine learning algorithms, pp. 27–40. Springer Nature (2022). https://doi.org/10.1007/978-3-031-21385-4_3
5. Bassil, R., et al.: Security analysis and solution for thwarting cache poisoning attacks in the Domain Name System. In: 2012 19th International Conference Telecommunication ICT (2012)
6. Adiwal, S., Rajendran, B., Pushparaj, S.D., Sudarsan, S.D.: DNS Intrusion Detection (DID)—a SNORT-based solution to detect DNS Amplification and DNS Tunneling attacks. *Frankl. Open* **2**, 100010 (2023). <https://doi.org/10.1016/j.fraope.2023.100010>

7. Anomaly Behavior Analysis of DNS Protocol. *J. Internet Serv. Inform. Secur.* <https://jisis.org/article/jisis-2015-vol5-no4-05/69803/>. Accessed 26 Jun 2023
8. Maksutov, I.C., Alekseev, M.: Detection and prevention of DNS spoofing attacks. In: 2017 Siberian Symposium on Data Science and Engineering (SSDSE), pp. 84–87 (2017). <https://doi.org/10.1109/SSDSE.2017.8071970>
9. Wu, H., Dang, X., Zhang, L., Wang, L.: Kalman filter-based DNS cache poisoning attack-detection. In: 2015 IEEE International Conference on Automation Science and Engineering (CASE), pp. 1594–1600 (2015). <https://doi.org/10.1109/CoASE.2015.7294328>
10. Huang, C., Zhang, P., Sun, Y., Zhu, Y., Liu, Y.: SFDS: a self-feedback detection system for DNS hijacking based on multi-protocol cross validation. In: 2019 26th International Conference on Telecommunications (ICT), pp. 238–243 (2019). <https://doi.org/10.1109/ICT.2019.8798832>
11. Wang, Z., Yu, S., Rose, S.: an on-demand defense scheme against DNS cache poisoning attacks. *EAI Endorsed Trans. Secur. Saf.* **4**(14) (2018)
12. Moubayed, E.A., Shami, A.: Detecting DNS typo-squatting using ensemble-based feature selection and classification models. *IEEE Can. J. Electr. Comput. Eng.* **44**(4), 456–466 (2021). <https://doi.org/10.1109/ICJECE.2021.3072008>
13. Ma, T., Xu, C., Zhou, Z., Kuang, X., Zhong, L., Grieco, L.: Intelligent-driven adapting defense against the client-side DNS cache poisoning in the cloud, pp. 1–6 (2020). <https://doi.org/10.1109/GLOBECOM42002.2020.9322430>
14. Jayasingh, B.: DNS cache poisoning attack analysis and detection using packet header. *CVR J. Sci. Technol.* **12**, 108–112 (2017)
15. I.-I. J. for I. R. in S. and Technology. A Framework for Security of DNS using Cryptography. Accessed 23 Jun 2023
16. Location Based Model for Prevention DNS Spoofing. In: Proceedings of the 2020 International Conference on Internet Computing for Science and Engineering. <https://dl.acm.org/doi/abs/10.1145/3424311.3424329>. Accessed 11 May 2023
17. Hussain, M.A., Jin, H., Hussien, Z.A., Abduljabbar, Z.A., Abbdal, S.H., Ibrahim, A.: Enc-DNS-HTTP: utilising DNS infrastructure to secure web browsing. *Secur. Commun. Netw.* **2017**, 1–15 (2017). <https://doi.org/10.1155/2017/9479476>
18. Pingle, B., Mairaj, A., Javaid, A.: Real-World Man-in-the-Middle (MITM) attack implementation using open-source tools for instructional use, pp. 0192–0197 (2018). <https://doi.org/10.1109/EIT.2018.8500082>
19. How DNSSEC Works, Cloudflare. <https://www.cloudflare.com/learning/dns/dnssec/how-dns-sec-works/>. Accessed 10 July 2024
20. AI-driven Comprehensive Cybersecurity: Seceon aiSIEM & aiXDR cGuard, Seceon. <https://seceon.com/aisiem-cguard/>. Accessed 18 Nov 2024
21. Lyu, M., Gharakheili, H.H., Sivaraman, V.: A survey on DNS encryption: current development, malware misuse, and inference techniques. *ACM Comput. Surv.* **55**(8), 1–28 (2023). <https://doi.org/10.1145/3547331>
22. Sanjay, B.R., Shetty, D.P.: DNS amplification & DNS tunneling attacks simulation, detection and mitigation approaches. In: 2020 International Conference on Inventive Computation Technologies (ICICT), pp. 230–236 (2020). <https://doi.org/10.1109/ICICT48043.2020.9112413>

Artificial Intelligence-Enabled Communication Systems



Predicting Coronal Hole Activity: Key to Mitigating Space Weather Impacts

Tarek Alsolame^(✉), B. D. S. Aritra, Enock Niyonkuru, Stephen Antogiovanni,
and Chandranil Chakrabortii

Trinity College, 300 Summit Street, Hartford, CT 06106, USA
{tarek.alsolame,bds.aritra,enock.niyonkuru,stephen.antogiovanni,
nil.chakraborttii}@trincoll.edu

Abstract. Coronal holes significantly impact space weather, making understanding and predicting their activity crucial for mitigating effects on Earth's technological infrastructure. This paper presents a novel approach to forecasting short-term coronal hole activity using a multi-head CNN-LSTM model trained on Solar Dynamics Observatory (SDO) dataset tabular data and images. Our methodology involves preprocessing, training, and implementing a prediction system to predict the solar surface ratio covered by coronal holes. We compare our results with baseline models, including LSTM, GRU, and ARIMA-based models, demonstrating improved predictive performance. Sensitivity studies analyze parameter impacts on performance. Our work contributes to filling the gap in predictive modeling for coronal hole activity, providing valuable insights for space weather forecasting and preparedness. The RESNET-50 based LSTM-CNN model outperforms comparable models by 34% in MSE and 27% in MAE in predicting sunspot ratio, identifying optimal look back and predict ahead windows that further improve performance by up to 15%. Future work includes extending the model to longer-term predictions and incorporating additional features for enhanced accuracy.

Keywords: Coronal holes · Coronal hole activity prediction · Convolutional Neural Networks · Timeseries prediction

1 Introduction

Coronal holes [7] are expansive regions within the Sun's corona characterized by reduced density and temperature. They hold significant scientific interest due to their open magnetic fields, which extend into space. This allows high-speed streams of solar wind to interact with Earth's magnetosphere, potentially triggering geomagnetic storms. Although such storms are commonly associated with solar minimums, they can occur at any time [6]. Coronal hole activity affects Earth in several ways, primarily through the emission of faster solar wind streams interacting with the magnetosphere, causing geomagnetic storms. These storms, categorized by National Oceanic and Atmospheric Administration (NOAA) [27],

can disrupt satellite operations, navigation systems, and radio communications, degrading high-frequency radio communications and increasing auroral activity visible even at lower latitudes [22]. Forecasters closely monitor coronal holes to predict their effects on space weather, which is vital for maintaining the reliability of space-dependent systems. Despite risks like disruption of communication systems and induction of electrical currents in power lines, the storms produced by coronal holes are typically less severe than those from coronal mass ejections (CMEs) [11]. Enhanced auroral activity from coronal hole activity can be visually stunning but also indicates disturbance in Earth's magnetosphere. While generally harmless on Earth's surface, intense auroral displays can pose risks to astronauts and space satellites due to increased radiation exposure. Despite their lesser severity, the long-lasting effects of coronal hole activity warrant advanced forecasting to prepare for potential impacts on technological systems and infrastructure [6, 11, 27].

Therefore, predicting coronal hole activity [21] is crucial for mitigating the effects of space weather on various technologies and services. These holes emit high-speed solar wind streams that, when interacting with Earth's magnetosphere, can cause geomagnetic storms with diverse impacts. These storms range from degrading high-frequency radio communications vital for aviation and maritime navigation to affecting satellite operations and GPS systems due to ionospheric changes. Furthermore, geomagnetic storms induced by coronal holes can disrupt power grids, leading to voltage instability or outages [31]. Predicting such events enables advanced warnings and protective measures, ensuring the resilience of technological systems and infrastructure. Astronomic forecasters monitor coronal holes closely to anticipate the strength and timing of solar wind streams, enhancing their ability to forecast geomagnetic conditions over the following days, crucial for space weather preparedness and the continuous operation of space-reliant technologies and services [33].

While the influence of coronal holes on space weather is widely acknowledged, there remains a gap in the literature concerning the prediction of their size and activity. While previous studies have predominantly focused on detecting increased coronal hole activity, accurately predicting such activity continuously is crucial. The size and positioning of coronal holes directly impact the intensity and effects of solar wind streams on Earth [38]. However, limitations in previous research on coronal holes stem from observational uncertainties in determining their exact boundaries. Automated detection schemes like ASSA CH, CHARM, and CHIMERA [30] vary significantly in results due to diverse methods ranging from intensity-based thresholding to complex algorithmic classifications. Additionally, the dynamic nature of coronal hole boundaries, sites of constant interchange magnetic reconnection, makes them challenging to define uniquely. These challenges, coupled with the transient nature of coronal holes, make consistent monitoring complex. Addressing these limitations requires a cautious approach when using coronal hole boundary data for space weather predictions and modeling. Continued improvement in detection methods and a better understanding of the physical dynamics at coronal hole boundaries are

essential for advancing our knowledge in this area and improving space weather forecasting capabilities.

Recent research [14, 26, 33] on coronal hole activity has significantly advanced our understanding of coronal holes and their impact on space weather. SDO [29] was launched in 2010 to understand and record solar variations affecting Earth. It operates in a geosynchronous orbit, equipped with three scientific investigations: Atmospheric Imaging Assembly (AIA), Extreme Ultraviolet Variability Experiment (EVE), and Helioseismic and Magnetic Imager (HMI) [3]. SDO provides full-disk images of the Sun in various wavelengths, magnetic-field maps, and Doppler-velocity maps [34], offering unprecedented insights into solar dynamics and space weather.

In contrast to prior work that focuses on detecting abnormal coronal hole activity, the proposed novel approach focuses on predicting the size of coronal holes in the short term (1–7 days). We employ CHRONOS [13] to produce highlighted, masked versions of coronal holes, serving as ground truth. Then, we pre-process input images to compute the coronal hole coverage ratio on the solar surface at 1-hour intervals (recorded as tabular data). The combined data, consisting of images and tabular data, is then fed into a multi-head CNN-LSTM model. The CNN-LSTM architecture combines the strengths of CNNs and LSTMs. CNNs excel at extracting features from input data, while LSTMs are adept at processing sequential information. This hybrid approach allows for effective sequence prediction by leveraging the spatial hierarchy of CNNs and the temporal learning capabilities of LSTMs to model intricate patterns in the data. The proposed approach, powered by the RESNET-50 CNN architecture, processes images and integrates both data types to predict future coronal hole activity with improved precision and accuracy. We describe our methodology in detail later in the paper. We present a comprehensive evaluation of various models for predicting sunspot ratio. Our results demonstrate the superior performance of the LSTM-CNN model, achieving the lowest Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) compared to other models. Sensitivity studies reveal that a look back period of 60 and a predict ahead window of 24 strike a balance between performance and computational overhead highlighting the effectiveness of the proposed method in capturing complex temporal patterns and its potential for applications in forecasting tasks.

2 Related Work

The Solar Dynamics Observatory (SDO) dataset [14] offers an unparalleled resource for solar physics research, providing high-resolution, multi-wavelength imagery of the Sun with exceptional temporal and spatial detail. The complexity and vastness of the SDO dataset make it ideal for image processing and machine learning techniques, which can uncover hidden patterns, improve predictions, and deepen our understanding of solar dynamics. Multiple works have leveraged this dataset to advance our understanding of solar phenomena and

develop innovative machine learning applications in heliophysics, including solar flare prediction using support vector machines [2], deep learning approaches for estimating coronal loop properties [10], flare precursor detection methods [5], and super-resolution imaging of magnetograms [14].

Research on coronal hole activity has traditionally emphasized detection methodologies, employing techniques ranging from manual inspection to automated computer vision methods to identify these regions on the sun [6, 7, 33]. Image processing techniques have been widely employed to detect coronal holes, characterized by cooler and less dense regions, using data from the observatories. Although successful in identifying their presence, these methods often fall short in providing predictive insights into their future behavior or evolution [14, 16]. Recent advancements [4, 20] have shown promise in analyzing space weather data through deep learning approaches. Prior research utilized a introduced a deep learning method using the U-Net architecture to detect coronal holes [4, 17, 20], highlighting the potential of deep learning in this domain, however their work was limited to coronal hole segmentation, matching, and map classification.

Jarolim et al. [19] proposed the CHRONNOS model, a convolutional neural network (CNN) based technique that utilizes multi-channel data for precise coronal hole boundary detection [1], demonstrating the suitability of convolutional neural networks (CNNs) for real-time applications. Despite these improvements in detection accuracy, the focus has remained primarily on identification rather than prediction. CHRONNOS has been used in multiple works for various types of analysis on solar data such as solar irradiance variability on decadal to millennial timescales [9], revisit the solar forcing recommendations [12], etc. In this work, we utilize CHRONNOS to provide the ground truth information regarding coronal hole activity which we detail in a later section. Existing literature lacks robust predictive models for coronal holes, focusing mainly on detection rather than future behavior prediction. Our research introduces a novel approach using the LSTM architecture to forecast coronal hole activity over the short term period (1–7 days). This advancement addresses a critical gap in the field, offering valuable lead time for preparations against potential space weather events. Our work fills this void, promising enhanced predictive accuracy and positioning our research as a notable contribution to the field.

3 Proposed Approach

This section outlines our proposed approach, consisting of three key components: data preparation, model training, and the coronal hole prediction system. We describe the dataset, data preparation, model training process, and the end-to-end pipeline, including baseline models used for comparison.

3.1 Dataset

As mentioned earlier, we utilized the comprehensive dataset from NASA’s SDO covering the years 2011 to 2018 [23]. This dataset includes a rich collection of

solar images and related data, with the baseline timestamp-related data obtained from prior work [19]. The dataset contained a CSV file that detailed temporal information for the period from 2011 to 2018, ensuring precise alignment with the corresponding image data. The SDO dataset comprises image matrices captured by the Atmospheric Imaging Assembly (AIA) and the Helioseismic and Magnetic Imager (HMI). Specifically, we included the AIA wavelengths of 94 Å, 131 Å, 171 Å, 193 Å, 211 Å, 304 Å, 335 Å, 1600 Å, and 1700 Å, as well as the HMI magnetic field components b_x , b_y , and b_z . For each hour, the file provides congruent paths to the respective image matrices, organized to facilitate seamless integration with the temporal data. The dataset was downloaded from the MiniSDO Dataset website [23] using the `wget` command. The data for each year provided as a .tar file that required extraction to access the individual image matrices. This extensive dataset (69 GB), combining high-resolution solar images across multiple wavelengths with precise temporal data, forms the foundation of our analysis, enabling a comprehensive study of solar phenomena, particularly focusing on the activity of coronal holes over time.

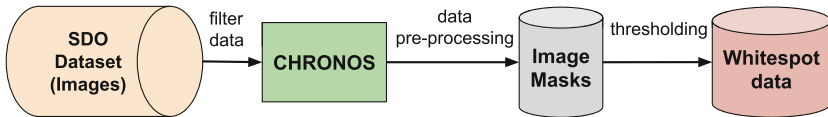


Fig. 1. Process for Computing sunspot Information.

3.2 Data Preprocessing

Following the acquisition of the dataset, we employed CHRONNOS [19] to pre-process the raw images, utilizing its convolutional neural network to accurately delineate coronal hole boundaries. This pipeline provided access to hourly image data from the SDO dataset, allowing for the generation of binary mask images via thresholding (mean pixel values of the mask image) [24]. Subsequent sunspot computation [32] involved counting sunspot ratios and total pixels, yielding the sunspot ratio (ratio of white pixels in a solar image divided by total pixels). We then merged this tabular data with the original CSV file, creating a comprehensive dataframe that included relevant AIA wavelengths and SDO image paths. The sunspot ratios (tabular data) were then normalized using min-max scaling to a range of 0 to 1, ensuring that all features contribute equally to the model's learning process.

The image data was retrieved using paths extracted from our tabular CSV file, which also contained the sunspot ratio information. We focused primarily on the 171 Å wavelength image data due to its minimal missing values compared to other wavelengths used to produce CHRONNOS mask outputs. We selected the 171 Å wavelength for our analysis due to both its availability and its established relevance in observing coronal holes. While we acknowledge that other

wavelengths contribute valuable information about solar phenomena, the 171 Å channel offers several advantages. It captures emission from the upper transition region and quiet corona, providing a clear contrast between coronal holes (which appear darker) and the surrounding regions. Moreover, the 171 Å channel is less susceptible to contamination from bright active regions and flares, which can obscure coronal hole boundaries in other wavelengths. The relatively low number of missing values in the 171 Å data further supports its suitability for our study, ensuring a consistent and reliable dataset for training and evaluating our model. We analyzed missing values by examining their occurrence at an hourly cadence, revealing varying levels of missing data across different years. Due to the inherent complexity of interpolating image data, we proceeded with training our multi-head machine learning model using the available data, acknowledging the potential for future refinements through tabular data interpolation techniques. In addition to the raw image data and the sunspot ratios, no other features were explicitly extracted or engineered for this study. The focus was on leveraging the spatial information from the images and the temporal patterns in the sunspot ratios to predict coronal hole activity.

3.3 Model Architecture

The model employs a multi-head architecture, comprising both CNN and LSTM components, implemented using the PyTorch [18] and TensorFlow [28] library. The CNN-LSTM architecture [25] integrates the strengths of CNNs for feature extraction from input data with the sequential processing capabilities of LSTM networks, enabling effective sequence prediction. Originally designed for visual time series prediction tasks, such as generating textual descriptions from image sequences, this hybrid architecture leverages the spatial hierarchy of CNNs and the temporal learning capabilities of LSTMs to model complex patterns in data.

Our CNN-LSTM model employed RESNET-50 as a feature extractor, processing input images to extract spatial features like shapes, textures, and patterns through convolutional and pooling layers. The extracted features were then fed into an LSTM network, capturing temporal dynamics and patterns, such as changes in features over time. Our CNN-LSTM model, by combining RESNET-50's spatial feature extraction and LSTM's temporal modeling, effectively learns complex patterns in spatiotemporal data. We modify the RESNET-50 architecture to accommodate 12-channel input, corresponding to the 12 wavelengths of SDO images used per time step. The final fully connected layer of RESNET-50 is replaced with a linear layer that outputs 128 features per time step, followed by a ReLU activation function. A two-layer bidirectional LSTM with 500 hidden units processes the CNN embeddings sequentially, with a dropout layer with a probability of 0.25 applied after each LSTM layer to prevent overfitting. This ensures that the model generalizes well to unseen data. The bidirectional structure captures temporal dependencies in both forward and backward directions. A final dense layer with ReLU activation and linear output maps the LSTM's output to the prediction horizon, generating forecasts for multiple future time

steps. For instance, if predicting 10 future time steps, the dense layer's output dimension would be 10. The model architecture is described in Fig. 2.

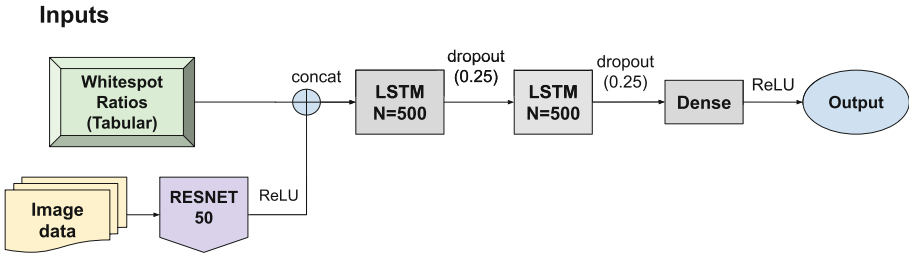


Fig. 2. Integrated Model Structure.

3.4 Model Training

The inputs to our CNN-LSTM model include both tabular and image data (represented as matrices). Due to computational resource constraints, we employed a batched training approach, partitioning the dataset into yearly chunks and training the model on each year's data separately. This strategy enabled efficient resource utilization and allowed the model to adapt to distinct yearly patterns. Given the yearly variations in our data, limited computational resources, and temporal shifts in data distribution, this approach proved effective. Within each annual batch, we utilized a fixed batch size of 64, further dividing each year's data into smaller, manageable batches. This hybrid approach combined the benefits of annual batching, such as adaptability to yearly changes, with the control and stability offered by fixed batch sizes, enabling optimal training performance.

To mitigate the risk of overfitting, we employ several regularization techniques during training, including dropout layers with a probability of 0.25 after each LSTM layer and early stopping, monitoring the validation loss and terminating training if it fails to improve for a specified number of epochs. This helps to prevent the model from memorizing the training data and ensures that it generalizes well to unseen examples. The effectiveness of these regularization techniques is evident in the fact that our CNN-LSTM model did not trigger early stopping during training, suggesting that it has learned meaningful patterns in the data rather than overfitting to noise. We utilize the Huber loss function with $\delta=1.0$ for training, striking a balance between Mean Squared Error (MSE) for small errors and Mean Absolute Error (MAE) for larger errors. To optimize the learning rate, we utilize a ReduceLR scheduler [15] with a minimum learning rate threshold, reducing the learning rate if the validation loss plateaus. We utilize the Adam optimizer [37] and early stopping with a patience of 10 epochs and monitoring of additional metrics (e.g., validation MAE) prevents overfitting.

The training data was reshaped to include a lookback and a predict-ahead parameter, which were adjusted during training for a comparative study. The

lookback parameter indicated the recent history of number of hours of sunspot and image data used to train the LSTM cells used as context (12–84, in intervals of 12). The predict-ahead parameter, on the other hand, specified the number of hours the model was predicting into the future, with ranges of 24 to 168 in intervals of 24 (1–7 days). The model was trained on a high-performance computing infrastructure that included an Intel Xeon server (16 cores, 1.6 GHz, 96 GB DRAM) and an NVIDIA Tesla RTX 4090 GPU for accelerated training. For evaluation, we employed a temporal split of the dataset, allocating the first 80% of I/O accesses (ordered by timestamp) for training and the remaining 20% for testing. We evaluate the model’s performance by comparing its predictions to the actual sunspot ratio, assessing its ability to capture the temporal dynamics of solar activity and generate accurate forecasts. Further analyses include comparing the model’s RMSE to a baseline RMSE obtained by predicting the average target value, as well as examining the minimum and maximum slopes of the predictions to understand the model’s sensitivity to changes in solar activity. This model architecture, combining the spatial feature extraction capabilities of CNNs with the temporal sequence learning capabilities of LSTMs, was designed to effectively capture and predict the dynamic patterns of coronal hole activity over time. By training the model in this manner, we aimed to achieve a high level of accuracy and reliability in our predictions, contributing valuable insights to the study of solar phenomena.

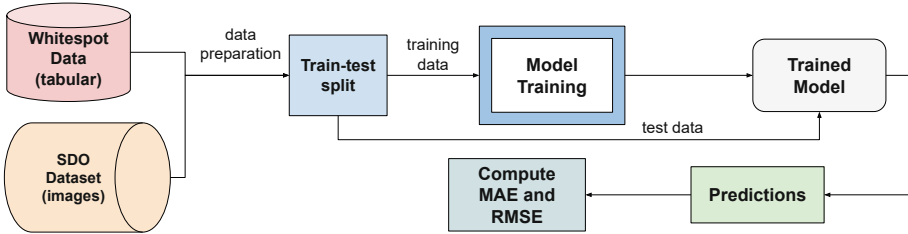


Fig. 3. Methodology

3.5 Coronal Hole Prediction System

Figure 3 illustrates the end-to-end coronal hole activity prediction system. Initially, raw image data from the SDO dataset was pre-processed to generate masked representations of solar images at different wavelengths using the CHRONNOS software. This process involved delineating the coronal hole boundaries and creating black-and-white mask images. We then computed the fraction of each image covered by coronal holes using a thresholding method [24] (Fig. 1). The resulting tabular data, which included the coronal hole coverage ratios, was combined with the image data collected at 171 Å. This combined dataset was used to data preparation for training the CNN-LSTM multi-head prediction

model, which forecasts short-term coronal hole activity over a range of 1 to 7 days. We train the model offline using 80% of the data (training data), enabling it to make predictions on the remaining 20% (test data) within the system. To evaluate performance, we utilize Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Squared Error (MSE) metrics, which provide a comprehensive understanding of the model’s accuracy and error distribution, allowing for comparison and benchmarking, and offering insights into strengths and weaknesses to guide future improvements. These metrics are well-suited for assessing the model’s ability to predict coronal hole activity and identifying areas for refinement.

The model leverages RESNET-50 for spatial feature extraction and LSTM networks for capturing temporal dynamics, effectively integrating both spatial and temporal information to enhance predictive accuracy. To evaluate the impact of different lookback periods and forecasting horizons, we conducted a sensitivity and ablation study. This analysis involved varying the lookback periods (in 12 h intervals) and forecasting horizons (in 24 h intervals) to assess the robustness and reliability of our model. For evaluation, we compared the forecasted sun spot value with the ground truth in the tabular data obtained from CHRON-NOS. Through this rigorous pre-processing and modeling approach, our system provides a powerful tool for predicting coronal hole activity, offering valuable insights for understanding solar dynamics and improving space weather forecasting.

3.6 Baselines

To establish robust baselines for evaluating our proposed CNN-LSTM model, we trained three distinct models: a standalone LSTM [35] model, which solely utilized the tabular data representing the fractional coverage of coronal holes on the sun’s surface to forecast future coronal hole activity; a GRU [8] model; and an ARIMA-based model, renowned for its efficacy in modeling and predicting time-series data [36]. The LSTM baseline allowed us to assess the contribution of temporal dependencies in the tabular data, while the ARIMA baseline provided a comparison with a traditional statistical approach, widely employed in time-series forecasting. By comparing our proposed model’s performance against these baselines, we can comprehensively evaluate its effectiveness in capturing complex coronal hole dynamics.

The selection of LSTM, GRU, and ARIMA as baseline models is grounded in their established efficacy and widespread use in time-series forecasting tasks. The LSTM and GRU models, representing recurrent neural network architectures, are particularly well-suited for capturing temporal dependencies in sequential data, making them relevant for modeling the evolution of coronal hole activity over time. The ARIMA model, a classical statistical approach, provides a robust benchmark rooted in traditional time-series analysis techniques. Furthermore, these models have been extensively employed in solar physics research, offering a direct comparison with existing approaches in the field. While acknowledging the emergence of more recent deep learning architectures, the chosen baselines

provide a solid foundation for evaluating the performance of our proposed model against established and widely recognized methods in both time-series forecasting and solar physics domains.

4 Results

In the following section, we present a comprehensive evaluation of the predictive performance of various models utilized in this study. The models under consideration include CNN-LSTM, LSTM, GRU, and ARIMA, each assessed based on their ability to predict sunspot ratio. We use Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) to evaluate model performance. RMSE is more interpretable than MSE, while MAE is less sensitive to outliers. MAE offers a robust measure, less sensitive to outliers. These complementary metrics provide a comprehensive understanding of model accuracy, robustness, and reliability.

Table 1. Results for different models (all years)

Model	MSE	RMSE	MAE	Early Stopped
CNN LSTM	4.22×10^{-4}	2.05×10^{-2}	2.18×10^{-2}	False
LSTM	1.21×10^{-3}	3.48×10^{-2}	2.60×10^{-2}	True
GRU	1.24×10^{-3}	3.52×10^{-2}	2.58×10^{-2}	True
ARIMA	1.53×10^{-3}	3.91×10^{-2}	2.85×10^{-2}	NA

4.1 Comparison with Baselines

The performance comparison of different models is summarized in Table 1. All models were trained using the initial 80% of the data from all years (2011–2018), while the remaining 20% was utilized for validation. The optimal values for the look back period and predict ahead window were determined using a grid search approach, where different combinations of these hyperparameters were evaluated based on their performance on the validation set.

The RESNET-50 based CNN-LSTM model demonstrates superior performance in predicting sunspot ratio, consistently outperforming other models across all evaluated metrics (MSE, RMSE, MAE). It achieves a 45% lower RMSE and a 42% lower MAE compared to the LSTM and GRU models. The improvement over the ARIMA model is even more substantial, with a 56% lower RMSE and a 58% lower MAE. This superior performance is particularly noticeable in years with high data variability, where the CNN-LSTM model maintains its accuracy while other models struggle. For the experiments, we identified the optimal configuration by setting the look back period to 60 h and the predict ahead

period to 24 h. This combination yielded the best performance, as demonstrated by our hyperparameter sensitivity study (shown later).

The CNN-LSTM model, leveraging both CNNs for spatial feature extraction and LSTMs for temporal dependencies, effectively handles the complex nature of sunspot ratio data. In contrast, LSTM and GRU models, while proficient in sequence modeling, show higher prediction errors, especially in years with significant fluctuations. Both also triggered early stopping, suggesting potential overfitting. The ARIMA model demonstrates the highest prediction errors, highlighting its limitations in capturing complex patterns. Due to computational constraints, with each epoch requiring over 3 h of training, we limited the number of training epochs to 30. We also implemented early stopping, halting training if validation loss failed to improve by at least 10^{-5} for five consecutive epochs. Notably, the CNN-LSTM model did not trigger early stopping, indicating potential for further improvement with additional training or hyperparameter tuning. In contrast, the LSTM and GRU models may benefit from regularization or architectural modifications to enhance their performance. Our results underscore the significance of model selection and architecture in achieving accurate sunspot ratio predictions. The CNN-LSTM model stands out as the most promising candidate for further development and real-world applications, particularly in scenarios characterized by high data variability.

Table 2. Performance Comparison of Different Models Over the Years

Year	RMSE				MAE			
	Our Model	LSTM	GRU	ARIMA	Our Model	LSTM	GRU	ARIMA
2011	1.148e-01	2.084e-01	2.087e-01	2.145e-01	8.223e-02	1.568e-01	1.591e-01	1.605e-01
2012	2.403e-02	3.615e-02	2.873e-02	3.000e-02	1.387e-02	3.076e-02	2.237e-02	2.292e-02
2013	5.242e-02	1.009e-01	1.209e-01	1.036e-01	3.418e-02	7.113e-02	1.002e-01	7.209e-02
2014	1.027e-01	1.820e-01	1.806e-01	1.926e-01	6.983e-02	1.350e-01	1.341e-01	1.451e-01
2015	4.663e-02	8.005e-02	7.929e-02	8.147e-02	2.664e-02	5.940e-02	5.831e-02	6.005e-02
2016	4.753e-02	9.247e-02	9.288e-02	9.950e-02	3.248e-02	6.942e-02	6.946e-02	7.387e-02
2017	1.056e-01	2.015e-01	2.008e-01	2.130e-01	7.727e-02	1.545e-01	1.526e-01	1.650e-01
2018	8.986e-02	2.288e-01	2.245e-01	2.158e-01	6.585e-02	1.469e-01	1.478e-01	1.436e-01

4.2 Year-Wise Performance Comparison

Table 2 presents a comprehensive comparison of our RESNET-50 CNN-LSTM model’s performance, trained on 80% of the data for each year, against baseline models LSTM, GRU, and ARIMA. The results unequivocally demonstrate the RESNET-50 based model’s superiority, consistently outperforming all baselines across all years in both RMSE and MAE metrics. This highlights the

model’s exceptional predictive capability, showcasing its potential for accurate forecasting.

In particular, our model demonstrated a significant improvement over the baselines in years with high data variability, such as 2011, 2013, 2014, and 2017. In these years, the RMSE of the RESNET-50 based model was 42–53% lower than that of the LSTM and GRU models, and 49–53% lower than that of the ARIMA model. For instance, in 2011, RESNET-50 achieves an RMSE of 1.148×10^{-1} compared to the Base LSTM’s 2.084×10^{-1} , Base GRU’s 2.087×10^{-1} , and ARIMA’s 2.145×10^{-1} . Similarly, in terms of MAE, RESNET-50 records 8.223×10^{-2} , while Base LSTM, Base GRU, and ARIMA show significantly higher errors of 1.568×10^{-1} , 1.591×10^{-1} , and 1.605×10^{-1} , respectively. These trends persist across subsequent years, highlighting the superior performance of the RESNET-50 LSTM-CNN model.

Table 3. Early Stopping Performance during Model Training

Year	Early Stopped		
	Our Model	LSTM	GRU
2011	False	True	True
2012	False	True	False
2013	False	False	True
2014	False	True	True
2015	False	True	True
2016	False	True	True
2017	False	True	True
2018	False	True	True

Furthermore, the early stopping criterion, implemented to prevent overfitting, was not triggered for the RESNET-50 based model in any of the years (Table 3), indicating that the model could potentially benefit from additional training. Its combination of convolutional and recurrent layers enabled effective capture of complex patterns, resulting in accurate predictions. In contrast, LSTM and GRU models frequently triggered early stopping, indicating higher susceptibility to overfitting and limited ability to learn complex patterns.

4.3 Sensitivity Studies

This section examines the impact of two key hyperparameters, look back and predict ahead, on the RESNET-50 based CNN-LSTM model’s performance. Due to computational constraints, we focus on data from 2018, enabling efficient experimentation. We analyze the trade-offs between forecast accuracy, and training complexity, providing insights for optimizing model configuration.

4.4 Impact of Look Back

The look back period, determining the number of previous time steps considered by the model, significantly influences the model’s performance. As can be seen in Table 4, increasing the look back period generally improves forecasting accuracy, with significant reductions in MSE, RMSE, and MAE scores. The most substantial improvement occurs between 48 and 60 h, indicating a sweet spot for balancing longer-term dependencies and computational complexity. The key findings are:

- Increasing the look back period from 12 to 60 h reduces MSE by 51.7%, RMSE by 31.0%, and MAE by 8.4%.
- The most significant improvement occurs between 48 and 60 h, with a 19.1% reduction in MSE, a 26.6% reduction in RMSE, and a 4.4% reduction in MAE.
- Beyond 60 h, the performance plateaus, indicating that incorporating more historical data does not lead to further improvements in forecasting accuracy.

These results suggest that a look back period of 60 h strikes a balance between capturing longer-term dependencies and avoiding excessive computational complexity. Increasing the look back period from 12 to 60 h reduces MSE by 51.7%, RMSE by 31.0%, and MAE by 8.4%. However, beyond 60 h, the performance gains diminish, and the model’s accuracy plateaus.

Table 4. Impact of Look Back on RESNET-50 LSTM-CNN Model Performance (2018)

Look back	MSE	RMSE	MAE
12	8.81×10^{-4}	2.97×10^{-2}	2.38×10^{-2}
24	7.61×10^{-4}	2.75×10^{-2}	2.17×10^{-2}
36	6.79×10^{-4}	2.60×10^{-2}	2.24×10^{-2}
48	5.22×10^{-4}	2.78×10^{-2}	2.28×10^{-2}
60	4.27×10^{-4}	2.09×10^{-2}	2.21×10^{-2}
72	5.21×10^{-4}	2.29×10^{-2}	2.51×10^{-2}
84	5.19×10^{-4}	2.28×10^{-2}	2.59×10^{-2}

4.5 Impact of Predict Ahead

The predict ahead parameter (PA), which determines the number of time steps into the future the model forecasts, significantly influences the performance of the RESNET-50 LSTM-CNN model. Table 5 presents the model’s performance for various predict ahead values in 2018. As the predict-ahead value increases, the model’s predictive accuracy generally decreases, as reflected by the increasing MSE, RMSE, and MAE values. This degradation in performance is expected, as

forecasting further into the future becomes inherently more challenging due to the accumulation of uncertainties and the potential for unforeseen events. Specifically, increasing the predict ahead from 24 to 48 h results in a relatively small increase in error metrics (RMSE increases by 8.3%, MAE by 9.2%). However, further increasing the predict ahead to 72, 96, 120, and 144 h leads to more substantial error increases (RMSE increases by 21.5%, 43.9%, 70.2%, and 107.3% respectively; MAE increases by 24.8%, 43.6%, 93.6%, and 117.4% respectively). These results indicate that the RESNET-50 LSTM-CNN model demonstrates good predictive capabilities for short-term forecasts (up to 48 h ahead), but its accuracy diminishes considerably for longer-term predictions. This finding underscores the importance of selecting an appropriate predict ahead value based on the specific application and the desired balance between accuracy and forecast horizon.

Table 5. Impact of Predict Ahead on RESNET-50 LSTM-CNN Model Performance (2018)

Predict Ahead	MSE	RMSE	MAE
24	4.25×10^{-4}	2.05×10^{-2}	2.18×10^{-2}
48	4.95×10^{-4}	2.22×10^{-2}	2.38×10^{-2}
72	5.25×10^{-4}	2.29×10^{-2}	2.72×10^{-2}
96	9.97×10^{-4}	2.64×10^{-2}	3.13×10^{-2}
120	8.96×10^{-4}	2.99×10^{-2}	4.23×10^{-2}
144	1.18×10^{-3}	3.40×10^{-2}	4.73×10^{-2}

5 Conclusion and Future Work

In conclusion, this paper presents a pioneering approach to forecasting short-term coronal hole activity, leveraging a multi-head LSTM model that integrates tabular data and image matrices from the SDO dataset. Our comprehensive methodology yields promising results in predicting coronal hole activity over a 1–7 day period, outperforming baseline models, including an LSTM model using only tabular data and an ARIMA-based model. This study addresses a critical knowledge gap in the field by focusing on coronal hole activity prediction, which is essential for mitigating its impacts on space weather-dependent technological systems and services. Through rigorous sensitivity and ablation studies, we demonstrate the robustness of our model and identify areas for future improvement. Our model’s promising results in predicting coronal hole dimensions are valuable for understanding space weather and its impact on Earth’s technological infrastructure. The size and evolution of coronal holes are directly linked to the properties of the solar wind streams they emit, which in turn can trigger geomagnetic storms and disrupt various technological systems.

Our work paves the way for future research directions, including extending the model to longer-term predictions, incorporating additional features, and exploring transfer learning applications. By advancing our predictive capabilities, we aim to contribute to developing more accurate and reliable space weather forecasting systems, ultimately enhancing the resilience of Earth's technological infrastructure against coronal hole activity impacts. Furthermore, we aim to establish a more explicit connection between our model's predictions and their physical implications, which could involve correlating predicted changes in coronal hole size with observed solar wind parameters, such as velocity and density, and investigating the relationship between these parameters and the occurrence and intensity of geomagnetic storms. By bridging this gap, we can enhance the practical utility of our model and contribute to the development of more effective space weather forecasting and mitigation strategies.

Acknowledgments. The authors wish to acknowledge the financial support provided by the NASA Space CT Grant Consortium under grant number Pro-Sum: P-2003. This funding enabled the completion of this research project and is greatly appreciated.

References

1. Abed, A.K., Qahwaji, R., Abed, A.: The automated prediction of solar flares from SDO images using deep learning. *Adv. Space Res.* **67**(8), 2544–2557 (2021)
2. Bobra, M.G., Couvidat, S.: Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm. *Astrophys. J.* **798**(2), 135 (2015)
3. Boerner, P., et al.: Initial calibration of the atmospheric imaging assembly (AIA) on the solar dynamics observatory (SDO). *Solar Dyn. Obs.* 41–66 (2012)
4. Chen, J., et al.: Ru-net: a residual u-net for automatic interplanetary coronal mass ejection detection. *Astrophys. J. Suppl. Ser.* **259**(1), 8 (2022)
5. Chen, Y., et al.: Identifying solar flare precursors using time series of SDO/HMI images and sharp parameters. *Space Weather* **17**(10), 1404–1426 (2019)
6. Choi, Y., et al.: Statistical analysis of the relationships among coronal holes, corotating interaction regions, and geomagnetic storms. *Sol. Phys.* **254**, 311–323 (2009)
7. Cranmer, S.R.: Coronal holes. *Living Rev. Sol. Phys.* **6**(1), 3 (2009)
8. Dey, R., Salem, F.M.: Gate-variants of gated recurrent unit (GRU) neural networks. In: 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), pp. 1597–1600. IEEE (2017)
9. Egorova, T., et al.: Revised historical solar irradiance forcing. *Astron. Astrophys.* **615**, A85 (2018)
10. Faber, J.T.: Characterizing Ultraviolet Bursts in a Solar Coronal Hole using Machine Learning Techniques. Master's thesis (2022)
11. Forbes, T.: A review on the genesis of coronal mass ejections. *J. Geophys. Res. Space Phys.* **105**(A10), 23153–23165 (2000)
12. Funke, B., et al.: Towards the definition of a solar forcing dataset for CMIP7. *Geosci. Model Dev. Discuss.* **2023**, 1–14 (2023)
13. Funke, B., et al.: Towards the definition of a solar forcing dataset for CMIP7. *Geosci. Model Dev.* **17**(3), 1217–1227 (2024)

14. Galvez, R., et al.: A machine-learning data set prepared from the NASA solar dynamics observatory mission. *Astrophys. J. Suppl. Ser.* **242**(1), 7 (2019)
15. Hammoud, M., Sakr, M.F.: Locality-aware reduce task scheduling for mapreduce. In: 2011 IEEE Third International Conference on Cloud Computing Technology and Science, pp. 570–576. IEEE (2011)
16. Hurlburt, N., et al.: Heliophysics event knowledgebase for the solar dynamics observatory (SDO) and beyond. *Solar Dyn. Obs.* 67–78 (2012)
17. Illarionov, E.A., Tlatov, A.G.: Segmentation of coronal holes in solar disc images with a convolutional neural network. *Mon. Not. R. Astron. Soc.* **481**(4), 5014–5021 (2018)
18. Imambi, S., Prakash, K.B., Kanagachidambaresan, G.: Pytorch. In: Programming with TensorFlow: Solution for Edge Computing Applications, pp. 87–104 (2021)
19. Jarolim, R., et al.: Multi-channel coronal hole detection with convolutional neural networks. *Astron. Astrophys.* **652**, A13 (2021)
20. Jatla, V., Pattichis, M.S., Arge, C.N.: Image processing methods for coronal hole segmentation, matching, and map classification. *IEEE Trans. Image Process.* **29**, 1641–1653 (2019)
21. Joselyn, J., McIntosh, P.: Disappearing solar filaments: a useful predictor of geomagnetic activity. *J. Geophys. Res. Space Phys.* **86**(A6), 4555–4564 (1981)
22. Korte, M., Stolze, S.: Variations in mid-latitude auroral activity during the Holocene. *Archaeometry* **58**(1), 159–176 (2016)
23. Kucuk, A., Banda, J.M., Angryk, R.A.: A large-scale solar dynamics observatory image dataset for computer vision applications. *Sci. Data* **4**(1), 1–9 (2017)
24. Leung, C.K., Lam, F.: Maximum segmented image information thresholding. *Graph. Models Image Process.* **60**(1), 57–76 (1998)
25. Livieris, I.E., Pintelas, E., Pintelas, P.: A CNN-LSTM model for gold price time-series forecasting. *Neural Comput. Appl.* **32**, 17351–17360 (2020)
26. McGregor, S., Dhuri, D., Berea, A., Muñoz-Jaramillo, A.: Flarenet: a deep learning framework for solar phenomena prediction. In: Neural Information Processing Systems (NIPS) 2017 Workshop on Deep Learning for Physical Sciences (DLPS), Long Beach, CA, US (2017)
27. Neupert, W.M., Pizzo, V.: Solar coronal holes as sources of recurrent geomagnetic disturbances. *J. Geophys. Res.* **79**(25), 3701–3709 (1974)
28. Pang, B., Nijkamp, E., Wu, Y.N.: Deep learning with tensorflow: a review. *J. Educ. Behav. Stat.* **45**(2), 227–248 (2020)
29. Pesnell, W.D., Thompson, B.J., Chamberlin, P.: The Solar Dynamics Observatory (SDO). Springer (2012)
30. Reiss, M.A., et al.: The observational uncertainty of coronal hole boundaries in automated detection schemes. *Astrophys. J.* **913**(1), 28 (2021)
31. Rust, D.M.: Coronal disturbances and their terrestrial effects: tutorial lecture. *Space Sci. Rev.* **34**(1), 21–36 (1983)
32. Smith, K., Landes, P.E., Thollot, J., Myszkowski, K.: Apparent greyscale: a simple and fast conversion to perceptually accurate images and video. In: Computer Graphics Forum, vol. 27, pp. 193–200. Wiley Online Library (2008)
33. Vršnak, B., Temmer, M., Veronig, A.M.: Coronal holes and solar wind high-speed streams: I. forecasting the solar wind parameters. *Solar Phys.* **240**(2), 315–330 (2007)
34. Yeo, K.L., et al.: The dimmest state of the sun. *Geophys. Res. Lett.* **47**(19), e2020GL090243 (2020)
35. Yu, Y., Si, X., Hu, C., Zhang, J.: A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* **31**(7), 1235–1270 (2019)

36. Zhang, G.P.: Time series forecasting using a hybrid arima and neural network model. *Neurocomputing* **50**, 159–175 (2003)
37. Zhang, Z.: Improved adam optimizer for deep neural networks. In: 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), pp. 1–2. IEEE (2018)
38. Zirker, J.B.: Coronal holes and high-speed wind streams. *Rev. Geophys.* **15**(3), 257–269 (1977)



Building Mapping for Implementing an Object-Oriented Model of Educational Infrastructure Usage at Jean Lorougnon Guédé (UJLoG) University in Central-Western Côte d'Ivoire

N'guessan Patrice Akoguhi¹(✉), Kpangni Alex Jérémie Koua¹,
Egomli Stanislas Assohoun¹, N'dri Pascal Kouame²,
and Koffi Fernand Kouame³

¹ University Jean Lorougnon Guédé of Daloa, UFR Environment,
Daloa, Côte d'Ivoire

{patrice.akonguhi, jeremiekoua, stanislas.assohoun}@ujlg.edu.ci

² University Félix Houphouët-Boigny of Abidjan, University Research and
Application Center in Remote Sensing (CURAT), Abidjan, Côte d'Ivoire

³ Virtual University, Unit of Research and Digital Expertise (UREN),
Abidjan, Côte d'Ivoire

fernand.kouame@uvci.edu.ci

Abstract. This study focuses on updating existing maps and enhancing their application for more efficient infrastructure management at UJLoG. The process was carried out in two distinct phases. First, it involved updating the educational building map based on a 2020 map, created through object-oriented supervised classification of a Very High Resolution (VHR) satellite image from the GeoEYE-1 satellite. The second phase involved the proposal and design of an object-oriented model using the Unified Modeling Language (UML) to optimize map utilization. The results indicate the presence of fifty-six (56) classrooms dedicated to educational activities, including sixteen (16) newly acquired spaces. Additionally, the study's outcomes support three key decision-making processes: the development of an interactive online map of teaching locations accessible at any time, the ability to generate detailed statistics on the utilization of these teaching areas, and enhanced oversight and monitoring of the usage, occupancy, and dynamics of the educational buildings used for training and other academic activities. The overall contribution of this study aims to optimize the management of education at UJLoG by developing a spatial model for planning educational infrastructures. The use of the k-NN algorithm enables the updating of building maps, thereby improving the accuracy and relevance of the available data.

Keywords: Mapping · UML modelling · object-oriented
classification · UJLoG-Côte d'Ivoire

1 Introduction

In Côte d'Ivoire, higher education, particularly at the university level, has experienced significant growth over the past decades in terms of both the number of institutions and student enrollments. This expansion has played a fundamental role in national development by contributing to the formation of high-value human capital, essential for the functioning of both public and private enterprises [1, 2]. However, numerous studies [3–6] and recent observations have highlighted the challenges faced by university institutions. These challenges include the continuous growth of the student population, a shortage of teaching equipment, and, most notably, a lack of adequate infrastructure such as lecture halls, classrooms, and laboratories.

Jean Lorougnon Guédé University (UJLoG) in Daloa, established in 2012 [7], exemplifies these issues. One of the university's major concerns is the effective planning of academic programs and the rational use of teaching facilities. This situation has led to functional difficulties that hinder the smooth conduct of academic and administrative activities. Optimal management, therefore, requires an integrated and coordinated approach to support decision-makers in planning and utilizing infrastructure, from reservation requests to strategic decision-making. To achieve this, the development of a robust database and its preliminary modeling is crucial.

In this context, the establishment of a digital infrastructure management model relies on object-oriented spatial modeling using Unified Modeling Language (UML) [1, 8–10]. The use of spatial technologies and geographic information systems (GIS) has revolutionized the management and analysis of infrastructure and land use. Since the 1970s, satellite imagery, initiated with the first Landsat satellites, has provided precise data on terrestrial resources. Advances in high-resolution sensors have improved these approaches despite some limitations related to resolution and difficulties in interpreting complex urban variables [11].

More recently, progress in very high spatial resolution (VHSR) imagery, with satellites such as Sentinel, Pléiades, QuickBird, and GeoEye, has enabled monitoring of urbanization and improved sustainable urban planning. Research has demonstrated the utility of combining satellite imagery with machine learning for finer analysis of land-use changes. For instance, deep learning models have been developed to predict and model urban dynamics with high precision [12, 13]. In Côte d'Ivoire, while VHSR analysis often focuses on detailed mapping and vegetation studies, it offers promising perspectives for urban management and infrastructure planning, including for academic institutions like UJLoG [14–16].

Alongside advancements in spatial analysis, classical approaches to processing satellite images, based on pixel-by-pixel classification, have gradually given way to object-oriented methods. This approach groups pixels according to contextual criteria, enabling a more accurate interpretation of spectral and spatial data. Such advancements underscore the importance of using advanced technological tools to address universities' growing needs for infrastructure planning and management.

The primary objective of this study is to develop an education management system at UJLoG by proposing a spatial model for planning the use of educational infrastructures, integrating both key stakeholders and the infrastructures themselves. First, the mapping of educational buildings will be updated using the k-nearest neighbors (k-NN) classification algorithm. Subsequently, a digital mapping model will be proposed for planning the university's educational infrastructure.

2 Material and Method

2.1 Overview of Study Area

UJLoG is a higher education institution under the Ministry of Higher Education and Scientific Research (MESRS) of Côte d'Ivoire. It is located in the city of Daloa, the capital of the Haut Sassandra region in the central-western part of the country (see Fig. 1).

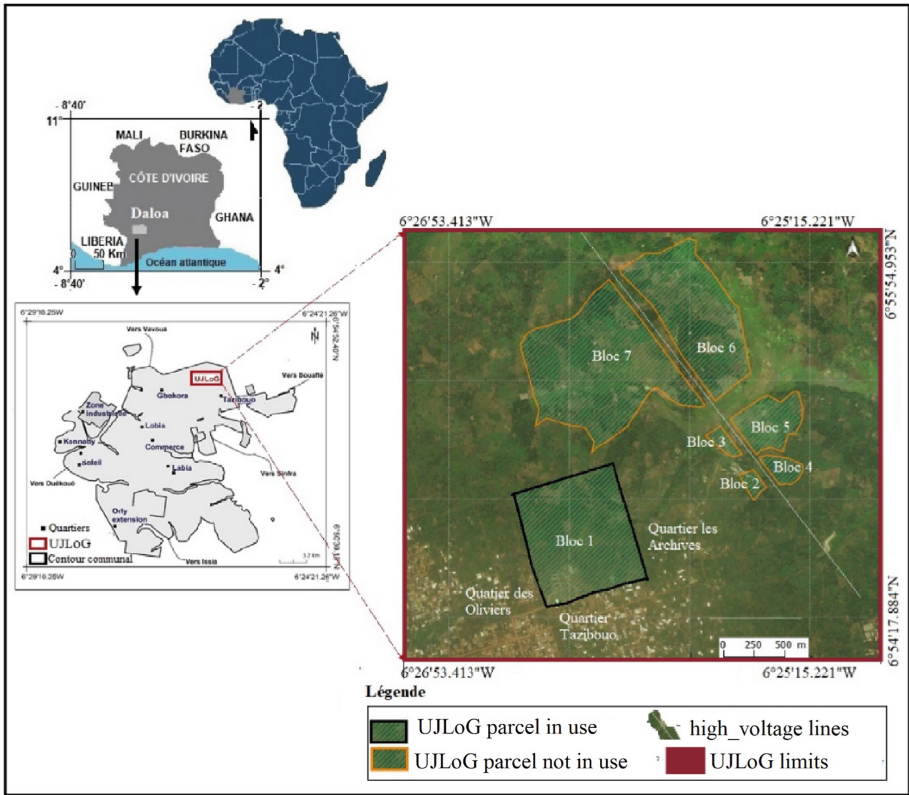


Fig. 1. Location of study area

The university is situated on a 301.16-hectare site, of which 87.3 hectares are currently in use. It is located in the northeastern part of the city, positioned between longitudes 6°26'53.413" and 6° 25' 15.221" West, and latitudes 6°54'17.884" and 6° 55' 54.953" North. UJLoG is bordered to the north by an expansion zone composed of vegetation, to the south by the "Tazibouo" neighborhood of Daloa, to the east by the "Archives" neighborhood, and to the west by the "Oliviers" neighborhood. As of 2024, this young university had an estimated student population of around five thousand (5,000) students [11]. The various academic programs are organized into five (5) faculties (UFR), which include Agroforestry, Environment, Economics and Management Sciences, Legal and Political Sciences, and Social and Human Sciences. Additionally, there is one (1) School of Continuing Education (EFC). The teaching is conducted by four hundred (400) faculty members, supported by one hundred thirty-eight (138) administrative and technical staff.

2.2 Materials and Data

The materials and data utilized in this study consist of cartographic data, satellite images, supplementary data, and field equipment.

Cartographic Data. To obtain a land-use map of UJLoG, it was necessary to use a 2023 Google Earth image for classification purposes.

Field and Data Processing Equipment. The field equipment includes a GPS device for recording the geographical coordinates of new buildings. The data processing equipment consists of a computer, an HP microcomputer (Intel Core i7 with 16 GB RAM and 256 GB SSD storage). This computer is equipped with the following software:

- QGIS 3.22 for updating the existing land use map
- PowerAMC 15, a graphical modeling tool used for creating various UML diagrams.

Additional Data. The additional data comprises curriculum templates from the different faculties, detailing the various programs (semester structure, type, content, and credit hours of the teaching units, etc.).

2.3 Methodological Approach

The general methodology is organized around three main stages. The first involves mapping the land states of UJLoG using object-oriented classification, a supervised classification method based on the k-nearest neighbors (K-NN) algorithm. The second focuses on implementing the parceling of teaching infrastructures at UJLoG. Finally, the last stage consists of modeling the spatiotemporal information related to teaching infrastructures.

Land Use Mapping (LUM) of UJLoG. The land use mapping, conducted through object-oriented classification, was carried out in three phases:

- Segmentation of the Google Earth image;
- Merging and creation of object shapes;
- Application of the algorithm and the actual extraction of objects.

Segmentation of the Google Earth Image. The image segmentation process divides the image into homogeneous segments based on criteria such as color and texture, thus creating distinct objects. A supervised classification approach, based on training areas manually selected through photo-interpretation, enables the identification of the target classes.

To enhance the classification's accuracy, segmentation scales ranging from 12% to 50% are used. Larger scales correspond to larger objects, while smaller scales are used for smaller objects (see Fig. 2).

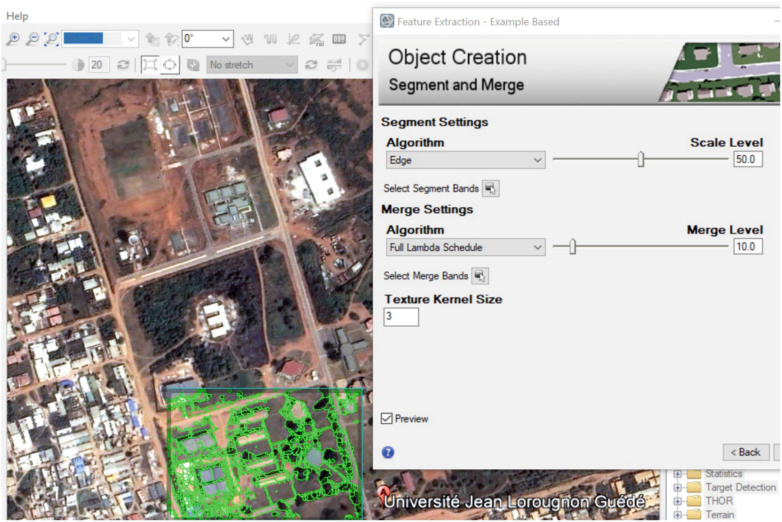


Fig. 2. Segmentation of the Image Using ENVI

Merging. Following the segmentation step, the merging phase combines similar segments to form larger entities, thus defining the total number of segments. A merging level of 15% was selected to account for both large and small objects. At the end of this process, the merged segments recreate the geographic shapes corresponding to the terrain.

Choice of Classification Algorithm. After completing the parameterization, the k-Nearest Neighbors (k-NN) algorithm was chosen for its proven effectiveness in shape recognition (Corbane et al., 2004). The k-NN is a classic supervised learning method that belongs to the category of Machine Learning algorithms. It is based on simple statistical principles and proximity rules to make predictions. Using sampling areas, the means (centroids) of each class are calculated. An unknown pixel is then assigned to the closest class, i.e., the one whose centroid is nearest to that pixel.

The flowchart in Fig. 3 provides a general summary of the methodology used to classify the UJLoG Google Earth image, aiming to produce the university’s land use map.

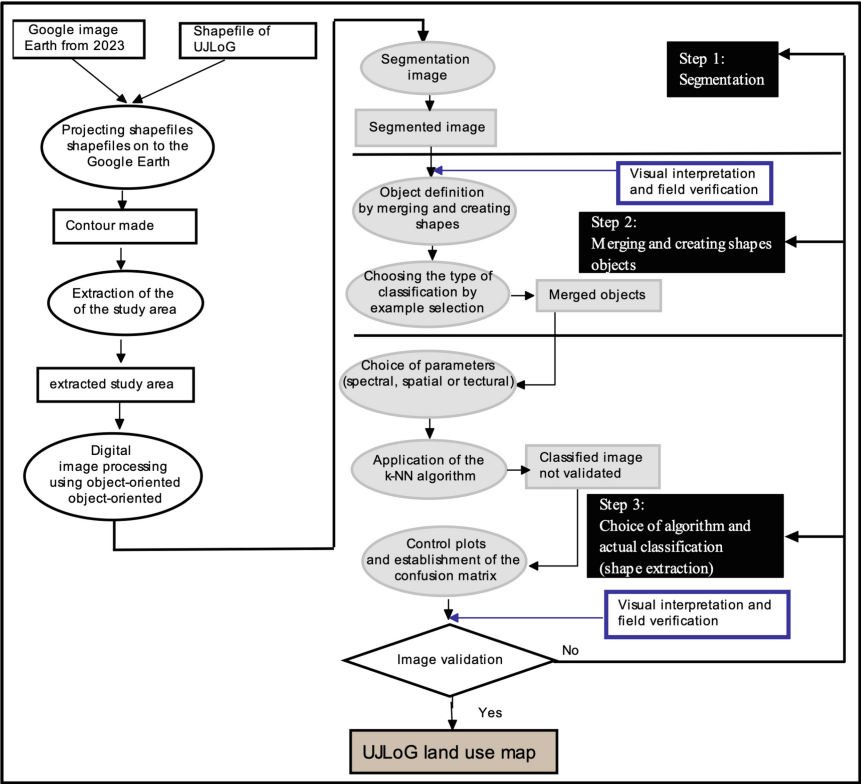


Fig. 3. Synthetic Flowchart of UJLoG Land Use Map Creation

The validation of the Google Earth image classification is based on three essential steps: field visits, visual interpretation of the classified image, and analysis using the confusion matrix. Field visits ensure a comparison between the classification results and real-world observations, guaranteeing accurate correspondence between satellite data and the observed reality. Visual interpretation

involves a detailed analysis of the morphological and spatial characteristics of objects to identify matches and discrepancies with their actual positions.

For this validation, three indices were used: global precision (PG), which indicates the proportion of correctly classified pixels; predicted precision (PP), which determines the accuracy of pixels within each specific category; and the Kappa index (K), which measures agreement [16]. These indices are computed using Eqs. 1, 2, and 3.

$$K = \frac{PG - PP}{1 - PP}, \quad (1)$$

$$PG = \frac{1}{N} \sum_{i=1}^{N_c} X_{ii}, \quad (2)$$

$$PP = \frac{1}{N^2} \sum_{i=1}^{N_c} (X_{+i} \cdot X_{i+}), \quad (3)$$

with :

- K : Kappa Index,
- PG : Global Precision,
- PP : Specific Precision,
- N : Total number of observations,
- N_c : Number of classes,
- X_{ii} : Number of observations in column i , and row i , located on the diagonal of the confusion matrix.
- X_{+i} : Total number of observations in column i ,
- X_{i+} : Total number of observations in row i .

Creation of the Teaching Infrastructure Parcels at UJLoG. The establishment of the parcel map was carried out in two main phases: the creation of LUM for buildings dedicated to teaching and the allocation of rooms through the parcel mapping.

Mapping of Land Use for Educational Buildings. The establishment of the parcel map was based on three key steps (see Fig. 4): (i) converting the classified and updated UJLoG image into a vector file supported by GIS software; (ii) exporting the obtained file into QGIS and displaying the various thematic classes as layers; and (iii) hiding layers (by unchecking them) that do not belong to the theme to be highlighted and validating the land use map for buildings dedicated to teaching purposes.

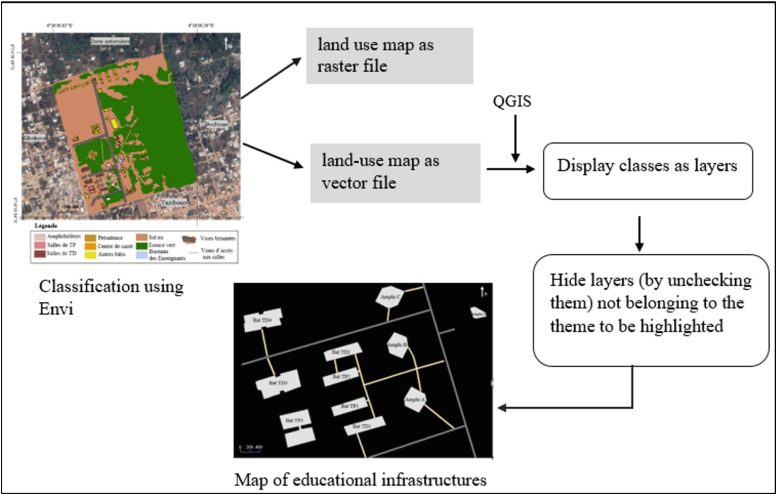


Fig. 4. Synthetic flowchart of UJLoG LUM Creation

Mapping Classrooms Through Parcels. This phase enables the creation of a room mapping and the proposal of a new coding system for buildings and rooms designated for teaching purposes (see Fig. 5). This process facilitates the localization of each building/room intended for teaching. It relies on two main steps: the first involves partitioning each polygon representing buildings into sub-polygons (rooms) based on their respective dimensions. The second step proposes a new coding system for buildings/rooms, following the recommendations of the UJLoG administration.

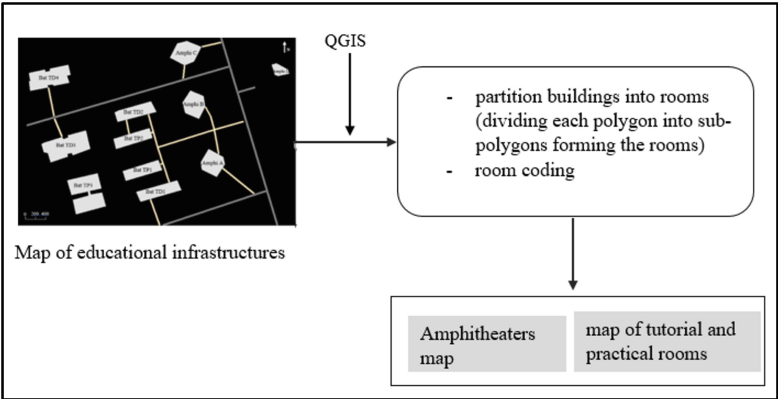


Fig. 5. Summary of parcel development and infrastructure codification methodology

Spatial Modeling Approach. The methodology used for modeling geomatic information systems has been the subject of several studies [12–15]. It is based on the Unified Modeling Language (UML). This methodology is divided into three essential steps, represented by different views: the functional model, the dynamic model, and the static model.

- **Functional Analysis:** The functional analysis illustrates the different use cases and actors. It relies on the use case diagram, which describes the system’s behavior from a user’s perspective in the form of actions and reactions.
- **Dynamic Analysis:** This step involves analyzing the evolution of objects over time using an activity diagram (AD).
- **Static Analysis:** The static view allows for the construction of the class diagram (CD), which is the most important aspect of object-oriented modeling. This part of the work focuses on the classes and their various associations.

3 Results

3.1 Land Use Map of UJLoG

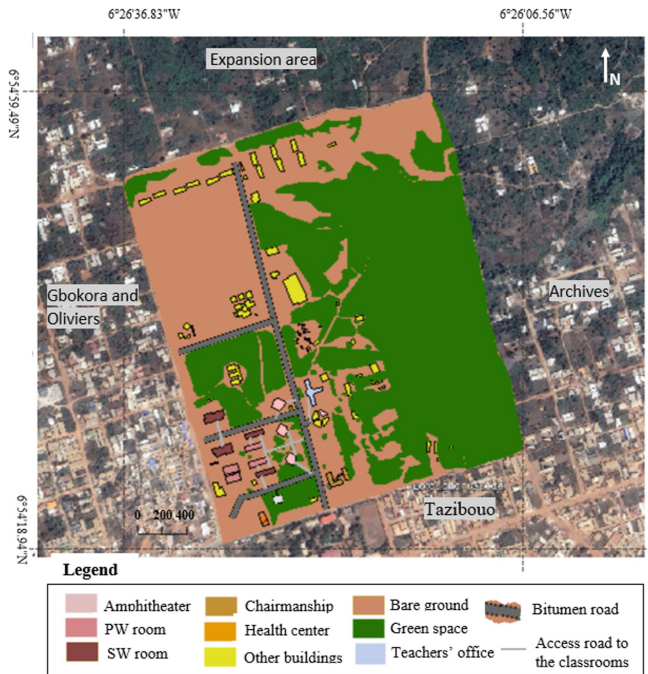


Fig. 6. Land use map of the UJLoG space of 2023

Figure 6 shows the land-use map obtained after classifying the 2023 Google Earth image (see Fig. 2). An initial visual analysis of the classified image reveals clear discrimination of urban objects. Field surveys identified five main land-use units across the university campus: “built-up area,” “green space,” “bare soil,” “paved road,” and “dirt road.”

The “built-up area” class includes low-density buildings used for teaching, research, administration, and other university-related activities. These structures, characterized by regular shapes, include both completed and under-construction buildings.

The visual analysis also highlights effective differentiation of other units. Green spaces and landscaped areas occupy a significant portion of the campus, interspersed with bare soil and tracks.

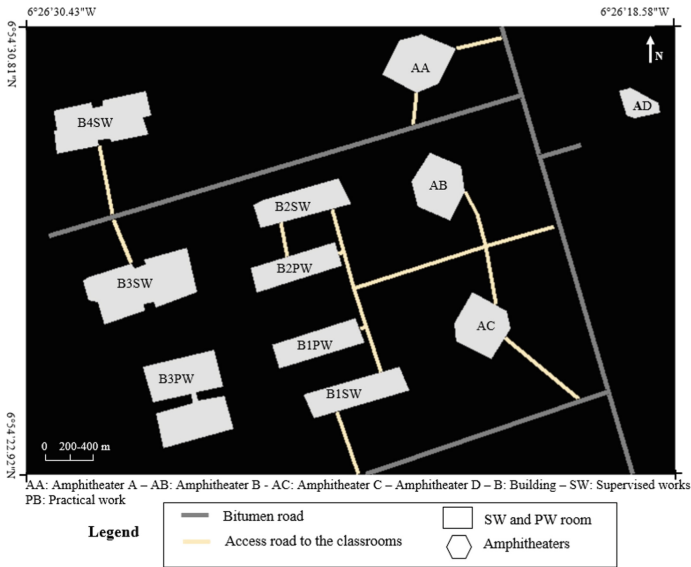


Fig. 7. Map of the supervise and practical rooms at UJLoG

3.2 Map of Teaching Infrastructure

The map in Fig. 7 illustrates the spatial distribution of buildings designated for various types of teaching activities at UJLoG. Eleven (11) buildings are represented on the map, namely: “Amphi AA,” “Amphi AB,” “Amphi AC,” “Amphi AD,” “B1SW,” “B2SW,” “B3SW,” “B4SW,” “B1TW,” “B2TW,” and “B3TW.” These buildings are categorized into two architectural styles.

The first type, consisting of eight (8) buildings, is represented by rectangular shapes that primarily denote facilities used for Tutorials (TD) and Practical

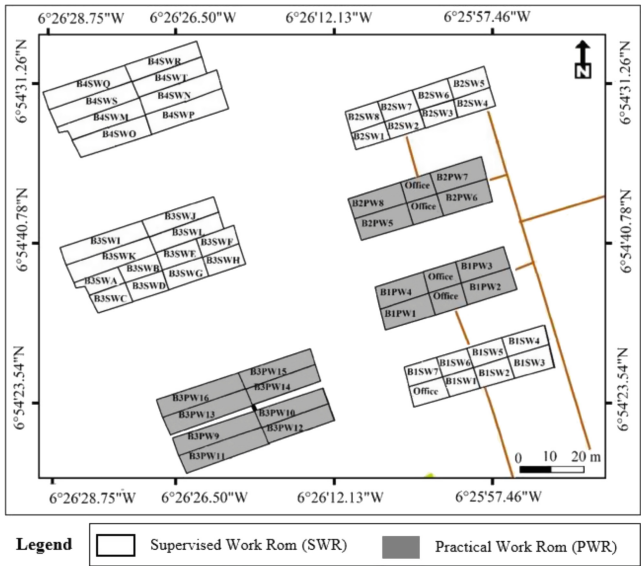


Fig. 8. Map of UJLoG supervised and practical room

Work (TP). The second type, symbolized by hexagonal polygons, includes four (4) buildings representing the amphitheaters.

A spatial analysis of the overall layout reveals a university area characterized by a relatively low density of buildings and vast natural spaces.

3.3 Nomenclature of Teaching Infrastructure

The coding of the buildings and classrooms has led to the establishment of a new nomenclature, as detailed in Table 1. The coding for the rooms designated for tutorials (TD) and practical work (TP) was modified to enhance clarity and integration into the model. The amphitheaters retain their original configuration and coding (Amphi A, B, C, D) for easier identification.

The spatial organization and nomenclature of the buildings and classrooms facilitated the creation of the university’s teaching infrastructure maps, as shown in Figs. 8 and 9.

3.4 Digital Model for Planning Educational Infrastructure

Functional View. The functional aspects of the system are illustrated by a general use case diagram, as shown in Fig. 10. The spatial analysis of UJLoG’s educational infrastructure is structured around use cases that integrate all the functions of the future database. The various actions performed within these use cases are categorized into “cartographic exploitation,” “planning consultation,” “teaching scheduling,” and “administrative tasks.”

Table 1. New nomenclature of buildings and rooms intended for teaching

Building coding	Rooms coding	Types of Courses	Number of buildings	Number of rooms
B1SW	B1SW1 to B1SW8	SW	1	7
B2SW	B2SW9 to B1SW16	SW	1	8
B3SW	B3SW _A to B3SW _L	SW	1	12
B4SW	B4SW _M to B4SW _T	SW	1	8
B1PW	B1PW1 to B1PW4	PW	1	4
B2TW	B2PW5 to B2PW8	PW	1	4
B3TW	B3PW9 to B3PW15	PW	1	8
Amphitheater	Amphitheater A to D	CM	4	4
Total			11	56

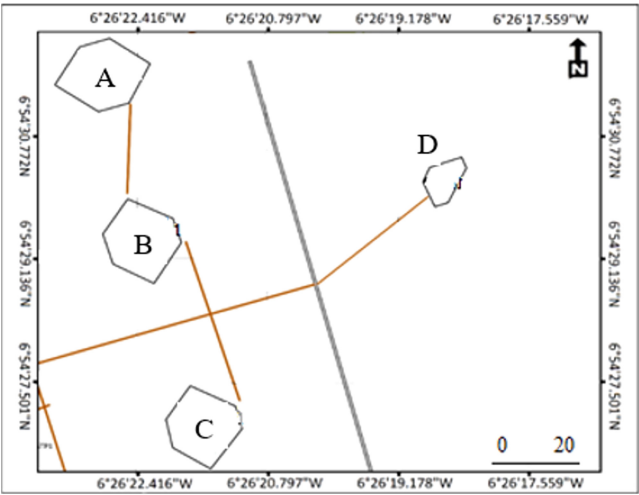


Fig. 9. Amphitheater map (Amphitheater A, B, C and D)

Dynamic View. The activity diagram (see Fig. 11) representing the dynamic view of this study facilitated the description of actions and the modeling of various interactive processes involved in room reservation management. This dynamic analysis also supported the identification of software objects, their attributes, operations, and associated methods. Additionally, it provided the foundation for constructing the class diagram.

Static View. The class diagram (CD) represents the model for the teaching infrastructure planning platform at UJLoG (see Fig. 12). This diagram consists of a set of polygons that define the classes, connected by polylines representing

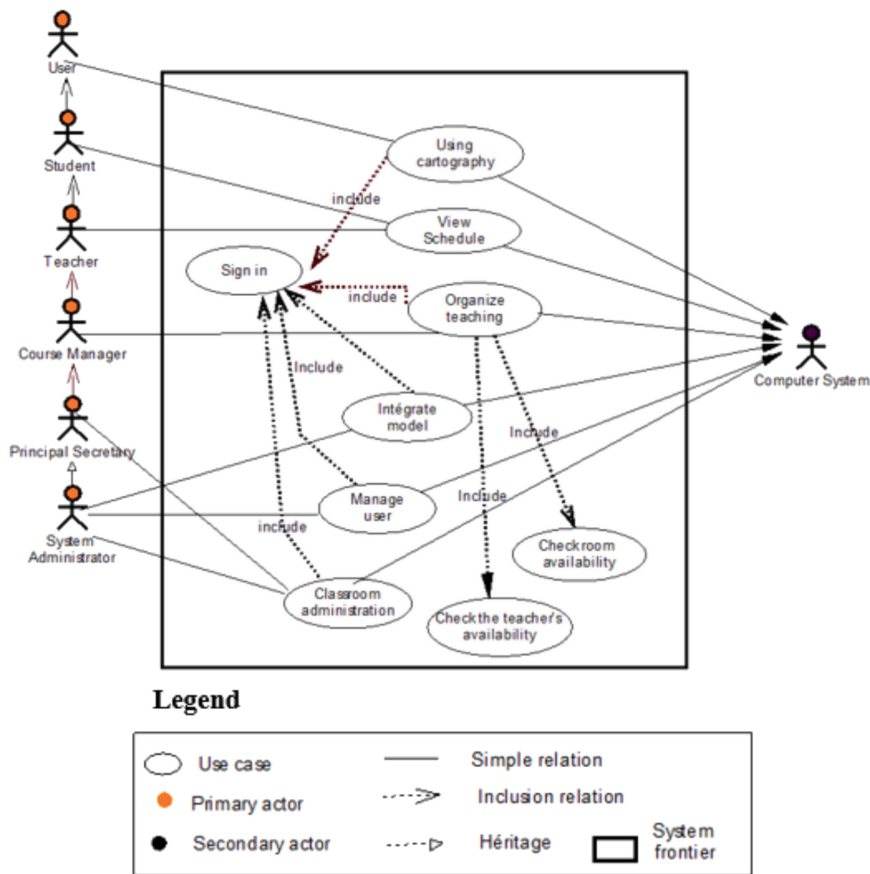


Fig. 10. Use case diagram named “general use”

semantic connections known as associations. It includes twenty-eight (28) classes, organized into spatial and non-spatial entities that are interconnected by various relationships.

The diagram outlines the selected entities and the expected processes, illustrating how the different components will be implemented within the database, encompassing both attribute data and spatial data. The teaching infrastructure planning model at UJLoG, as depicted by the CD, is a system composed of interacting objects. The various classes are interlinked through associations that establish relationships between them (Fig. 12).

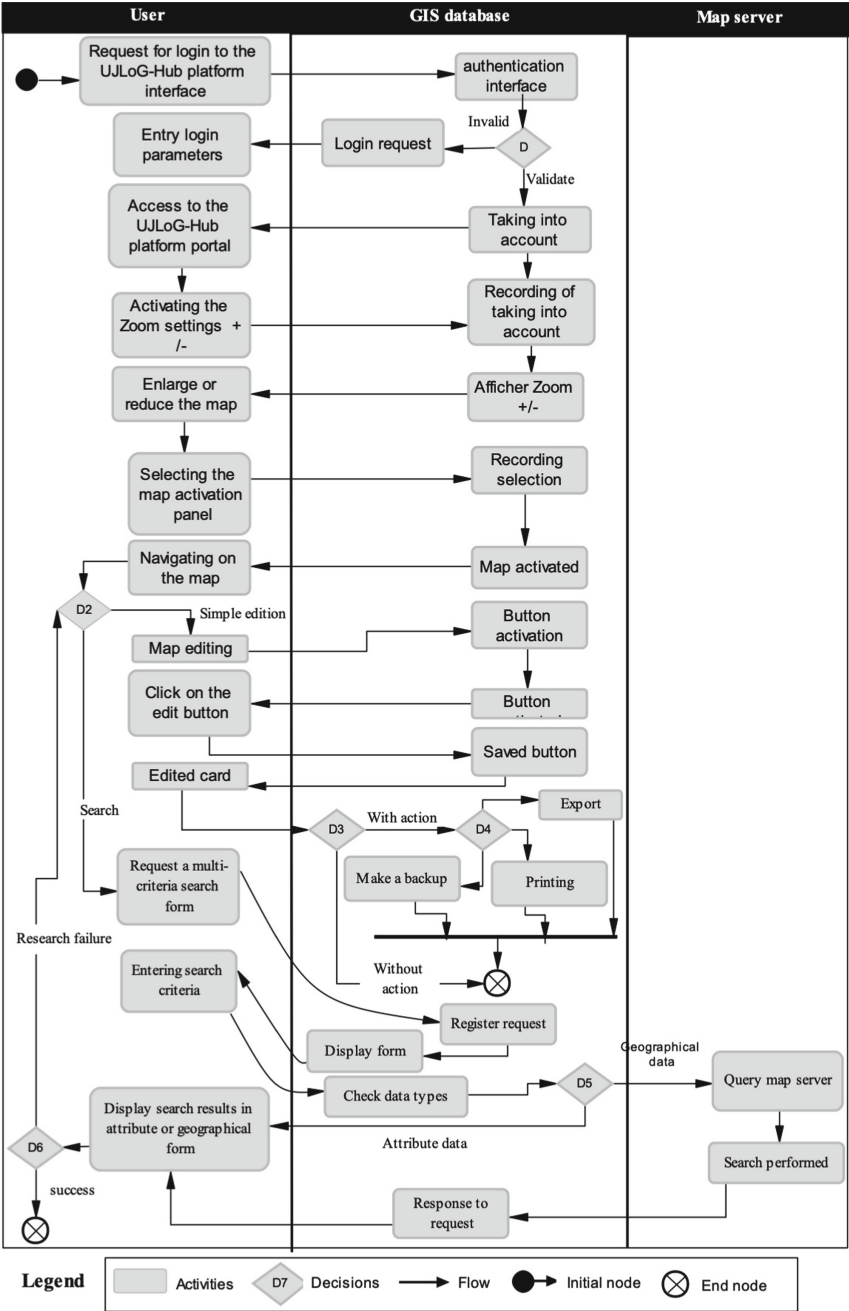


Fig. 11. Activity diagram illustrating the cartographic operating procedure of the UJLoG-Hub platform

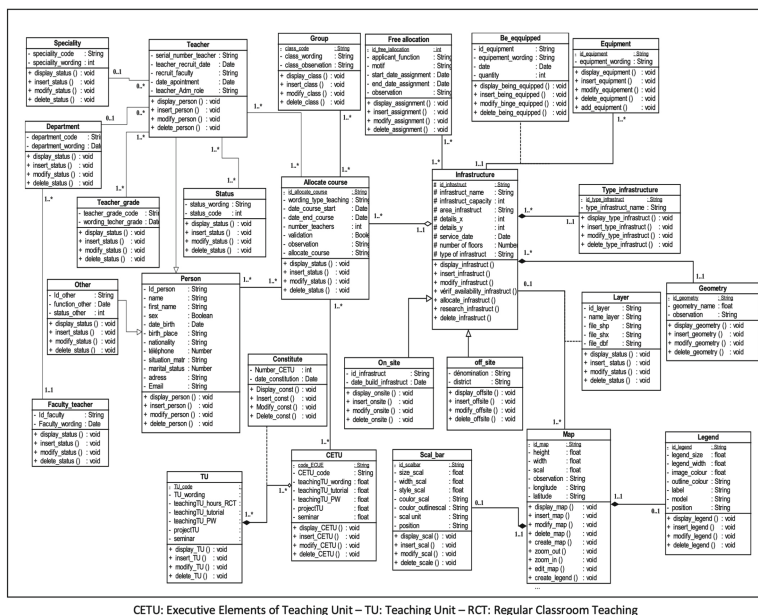


Fig. 12. Class Diagram representing the object-oriented model of educational infrastructures

4 Discussion

The modeling of educational infrastructure at UJLoG has enabled the development of a framework for scheduling teaching activities and monitoring the associated infrastructure. This work lays the groundwork for a future Spatial Database (SDB) that will manage and track the dynamics of teaching infrastructure and its utilization via the web. The study enhances access to spatial data, supports data visualization and exploration, and creates an adaptable tool for the analysis and processing of both geographic and descriptive data for management activities [16,22]. To this end, the model was developed using a rigorous methodological approach grounded in new internet-based technologies. From a technological and methodological perspective, the UML approach addresses the key phases of requirements analysis and object-oriented programming, as evidenced by several research studies [15,17]. Moreover, for some researchers [9,18,19], the adoption of an object-oriented approach to modeling is justified by its capacity to standardize object models, independent of implementation and deployment platforms (such as programming languages, work environments, operating systems, and network architectures).

This work, therefore, demonstrates the relevance of the key concepts of the UML for system modeling, particularly in the context of territorial management and analysis. Similar studies have also shown that UML, through the use of a class diagram (CD), effectively reveals the overall structure of the application

[20–25]. The CD has helped establish a more concise representation of the information system, defining the relationships between spatial data and descriptive data. It is important to note that the emphasis placed on the CD indicates that a high-quality model is crucial for ensuring the reliability of the database modeling that drives a platform [22, 25, 26]. According to the results illustrated by the model, it is evident that the modeling relies primarily on the object-oriented (OO) model. This approach translates the real world into objects by encapsulating data and various processes, thereby facilitating the understanding of the architecture and the manipulation of data. However, UML-based modeling has limitations when it comes to modeling complex geographic data. The methodological approach adopted seems to focus on a standardized design for geographic models, allowing for normalized representations of spatial data. This can be considered a limitation in the implementation of web-based geomatics applications that manage complex spatial objects such as sets of lines, points, and complex areas. It would be advisable to adapt this UML-based model for the modeling of spatial data.

5 Conclusion

This study produced two key outcomes. The first demonstrated the potential of Google Earth for object-oriented classification in updating land-use maps at UJLoG. The second utilized this updated data to propose an object-oriented model for educational infrastructure, aimed at improving the management of academic programs. The spatial model implemented for teaching planning effectively represents all entities (object classes) within the system and their relationships. Ultimately, the model, as represented by the class diagram, supports management at two levels: the interactive organization of infrastructure and the continuous monitoring of its temporal evolution and spatial distribution.

The results obtained will be valuable for the efficient digital management of classroom allocation and other educational activities. Furthermore, it would be beneficial, in future work, to enhance the UML model by incorporating more advanced spatio-temporal concepts from the MADS method (Modeling for Applications of Data with Spatio-temporal extensions). MADS, as a scenario-based modeling approach, focuses on representing geographical aspects by considering different use-case scenarios. Developed in the field of spatial systems analysis and management, MADS would thus be useful for modeling complex systems where space and time play a crucial role in future modeling efforts.

References

1. INRA: Modeling and agrosystems, all files on modeling and agrosystems (2018). <http://www.inra.fr/Chercheurs-etudiants/Systemes-agricoles/>
2. Kant, J.-D.: Apport des mathématiques et de l'informatique pour la modélisation en sciences humaines et sociales. *Math. Sci. Hum.* **197**, 47–64 (2012)

3. Traoré, G.: La crise de l'Université en Afrique: Diagnostic et éléments de stratégies transversales. In: Actes du Colloque International Pluridisciplinaire. Revue Ivoirienne de Philosophie et de Sciences Humaines, Volume XII, vol. 22, p. 32 (2022)
4. K. H. Amon, K. Tiemoman, Sehi Bi, A.M.: Étude d'un dispositif techno pédagogique en formation ouverte et à distance: le cas de l'Université Virtuelle de Côte d'Ivoire (UVCi), *Frantice.net*, vol. 17, p. 14 (2020)
5. CNDHCI: Sur la mise en œuvre du droit à l'Education en milieu universitaire: les difficultés du système universitaire de Côte d'Ivoire, Rapport de la commission Nationale de Droits de l'Homme de Côte d'Ivoire, p. 10 (2018)
6. Awoku, K.: Les Universités africaines francophones face à l'innovation: discours des apprenants sur le web dans l'enseignement supérieur au Togo (2012). <http://www.adjectif.net/spip/spip.php?article198>
7. UJLoG: Décret n°2012-981 du 10 Octobre 2012 déterminant l'organisation et le fonctionnement de l'Université Jean Lorougnon Guédé de Daloa, p. 24 (2012)
8. Frihida, A.: Contribution à la modélisation des données spatio-temporelles. Traitement du signal et de l'image. Thèse de Doctorat de l'Ecole Nationale d'Ingénieurs de Tunis (Université de Tunis El Manar), p. 133 (2025)
9. Terre.ESA: Rapport 2023 de l'ESA sur l'environnement spatial, the European Space Agency (2023). https://www.esa.int/Space_in_Member_States/France/Rapport_2023_de_l'ESA_sur_l'environnement_spatial
10. Gba, N.E., Saley, M.B., Kouassi, K.D.: Contribution of a WebSIG to a study of the distribution of education staff for vocational training in Cote d'Ivoire. *Sci. Res. Publ. Intell. Inf. Manag.* **13**, 141–155 (2021)
11. Kouakou, G.-C.D.: Distribution spatiale et dynamique de la population de palmiers rôniers, *Borassus aethiopum* Mart, par approche de la télédétection et du Système d'Information Géographique (SIG) de la réserve de Lamto (Centre de la Côte d'Ivoire), Thèse de Doctorat, Université Nangui Abrogoua, Abidjan (Côte d'Ivoire), p. 253 (2020)
12. Reynaud, M., Malet, J.-P., Déprez, A.: Approche orientée objet pour la détection automatique des glissements de terrain en zone de montagne: application à Haïti, Ecole et Observatoire des Sciences de la Terre (Université de Strasbourg), p. 10 (2020)
13. Affian: Apport de la géomatique dans l'adressage de l'Université Felix Houphouët-Boigny en vue de la réalisation d'un réseau intelligent, Master, UFHB, p. 62 (2016)
14. Dongo K., Kouame K.F., Kone B., Biemi J., Tanner M., Cisse G.: Analyse de la situation de l'environnement sanitaire des quartiers défavorisés dans le tissu urbain de Yopougon à Abidjan, Côte d'Ivoire. *Vertigo*, vol. 3, pp. 1–11 (2008)
15. Lhomme, S.: The development of geographic information on the Internet is linke to the development of applications on the web, Doctoral thesis, Paris Diderot University-France, p. 376 (2012)
16. Semlali, E.H., Azon, S., Elarchi, E.: Design and development of a GIS prototype for managing the monitoring of the execution highway work. In: Regional Conference, Marrakech—Morocco, p. 23 (2003)
17. Plateforme de cours en ligne-UJLoG à distance et organisation des enseignements (Côte d'Ivoire, Daloa) (2024). <https://ifoad-ujlog.net/master/>
18. Bisone, F., Etienne, L., Devogele, T.: Modeling and extraction of trajectory semantics from multi-sensor data, Lavoisier, Hal-02110261. *Int. J. Geomatics*, 461–483 (2018)
19. Khalfi, B.: Modeling and construction of fuzzy geographic databases and maintaining model consistency for SQL and NoSQL DBMS, Doctoral thesis, Paris Lumière University and Paris 8 University, p. 211 (2017)

20. Malek, B.Y.: Conceptual modeling for real-time geomatics applications, Doctoral thesis, INSA Lyon-France, p. 200 (2010)
21. Adous, F.G., Elayachi, M.A., El Ghouat, M.A., Ibnain, F.: Modélisation géospatiale de la gestion des aides agricoles dans le cadre du plan Maroc vert. *Rev. Mar. Sci. Agron. Vét.* **5**, 163–196 (2016)
22. Aye, Z.C.: A collaborative WEB-GIS based decision support platform for risk management of natural hazards, Doctoral thesis, University of Lausanne-Switzerland, p. 270 (2016)
23. Lonchamp, J.: Analyse des besoins pour le développement logiciel, Recueil et spécification, démarches itératives et agiles, Ed Dunod, p. 319 (2015)
24. Khalfi, C., De Runz, C., Faiz, S., Akdag, H.: improvement of the f-perceptory approach through management of fuzzy complex geographic objects. In: *Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume II /W5: ISPRS Geospatial Week 2015, 28 Sep–03 Oct 2015, p. 8. La Grande Motte, France (2015)
25. L. Auxepaules, Analyse des diagrammes de l'apprenant dans un EIAH de la modélisation orientée objet: le système OADC, thèse de Doctorat, Université du Mans (France), p. 234 (2009)
26. Konan-Waidhet, A.B.: Conception d'une base de données spatiales en vue d'une gestion intégrée des Ressources du territoire par la méthode Merise. In: *Cas de la région de Denguélé (Nord-Ouest de la Côte d'Ivoire)*. *Revue de Géographie Tropicale et d'Environnement*, vol 6, p. 14 (2007)



Optical RIS in Indoor VLC Environments: What are the Suitable Materials

Frank Igboamalu^{1,2}(✉)() , Alain Ndjiongue^{1,2}() , Khmaies Ouahada^{1,2}() ,
and Ling Cheng^{1,2}()

¹ Faculty of Engineering and the Built Environment, University of Johannesburg,
Johannesburg, South Africa
igbofrank@gmail.com

² Department of Electrical and Electronic Engineering Science,
Johannesburg, South Africa

Abstract. This paper explores the use of several available materials in optical reconfigurable intelligent surfaces (ORISs) for application in indoor visible light communication (VLC) environments. To achieve this, we determine the reflectivity of selected materials. We evaluate the corresponding ORIS-assisted indoor VLC channel gains and the achievable rates associated with these materials. Simulation results show that recycled materials, like can lids, lead to lower reflectivity, channel gains, and transmission rates. However, they are available worldwide and free, and their utilization in ORIS contributes to a clean living environment. The results obtained with a liquid crystal (LC)-based ORIS show that we can achieve more than 100% of reflectivity if a dye such as a concentration of trinitrofluorenone is added to the LC's substance and helps to mitigate the data loss imposed by the double fading due to the introduction of the ORIS module into the channel.

Keywords: Optical wireless communication systems · optical reconfigurable intelligent surfaces (ORISs) · ORIS materials · liquid crystals · recycled materials · non-line-of-sight channel gains · achievable rates

1 Introduction

Visible Light Communication (VLC) systems represent a significant advancement in wireless communication, leveraging the light spectrum to achieve high-speed data transmission [1–3]. The advent of Optical Reconfigurable Intelligent Surfaces (ORIS) further enhances VLC systems by improving coverage, resolving signal obstructions, and optimizing non-line-of-sight (NLoS) signal paths. These advancements align with the evolving demands of beyond-5G and 6G

Supported by University of Johannesburg.

networks, which require high data throughput and low latency. The number of research outputs in the literature on optical reconfigurable intelligent surfaces (ORIS)-assisted indoor VLC systems indicates that ORIS technology is getting mature and demonstrates validation of the concept [4–9, 12]. The following step is the mass production of ORIS modules for indoor VLC environments. In future implementations, this will be helpful beyond fifth-generation and sixth-generation mobile networks, which transmit a high amount of data between devices with low latency. To this end, utilizing suitable materials in the reflective planar units used in the ORIS module is necessary. The ORIS module is an optical phased array that convoys the input signal to a targeted area and arranges the transmission channel. It consists of electronically controllable elements and has low power consumption, which is advantageous when comparing the relay system and other competing technologies. This paper investigates the channel gains and transmission rates of the non-line-of-sight (NLoS) through the ORIS paths of ORIS-aided indoor VLC systems for several ORIS available materials.

In recent years, using ORIS in the indoor VLC environment has enhanced data transmission to solve the signal obstruction and skip-zones dilemmas modifying the transmission channel [4–9], thereby improving the capacity of the system. The VLC systems also help to enhance the receiver's field of view (FoV) [15, 16], with the introduction of the RIS technology to achieve a transmission rate. ORISs also help solve the line-of-sight (LoS) blockage dilemma by strengthening the NLoS signal, solving dead system zones within the RF network environment, and initiating an intelligent communication channel [13, 14]. In addition, ORIS technology helps to improve the signal-to-interference plus noise ratio (SINR) and the system's spectral efficiency. Furthermore, ORISs like any other variants of RIS, enable reconfigurable VLC environments. Most works and results proposed in the literature assume the reflectivity of ORISs. For example, the authors in [9] analyze the capacity of ORIS-assisted VLC systems with hybrid non-orthogonal multiple access upon the assumption of a 0.95 ORIS reflectivity. In [5], the authors exploit the ORIS technology to address the LoS signal blockage in indoor VLC systems assuming a 0.95 ORIS reflection coefficient. Similarly, the work in [10] considers that the reflectivity of the reflective elements is 0.9 in a study of ORIS-assisted multiple-input multiple-output VLC systems that combine the design of the surface configuration and transceiver signal. Studying the temporal characteristics of the ORIS-aided VLC channel using radiometric concepts is another example that assumes a reflection efficiency of 0.8 [11]. Although the work in [12] compares meta-surfaces to mirror arrays for ORISs, the simulation setup still assumes a reflectivity of 0.8 for both types of materials. This fundamental property of the ORISs (reflectivity) relates to the material used in the ORIS and undeniably affects the channel gain, transmission rate, and system performance. Among the key properties of ORIS materials is their reflectivity, which has a significant impact on system performance. Materials with higher reflectivity result in better channel gains, higher data rates, and improved spectral efficiency. While ORIS-assisted VLC systems have received

considerable attention in the literature, previous studies often assumed fixed reflectivity values without taking into consideration the material properties that may influence these values. Most existing literature focuses on theoretical performance metrics without examining the practical implications of material selection. This gap motivates the present study to evaluate a wide range of materials, including recycled options, for their reflectivity, environmental sustainability, and practical deployment potential.

We investigate several potential ORIS materials for indoor VLC environments. Based on their reflectivity, we analyze the channel gains and achievable rates of ORIS-assisted indoor VLC systems. We group these into three sets, including (i) sheet reflectors (mirror array) like polished anodized Aluminum (PAA), Mylar, Aluminum foil, acrylic mirrors, glass mirrors, and Astro-foil, (ii) liquid-based reflectors (meta-surface) such as liquid crystals (LCs) including A4907 and 3T-2MB combined with 0.01 wt% of trinitrofluorenone (TNF), and (iii) reflectors based on recycled materials (mirror array) like can lids. Even though recycled materials demonstrate low reflectivity, channel gains, and achievable rates, they are available worldwide for free, and their utilization in ORIS contributes to cleaning our living environments. With LC-based ORIS, it is possible to realize a reflectivity higher than 1, which helps to solve the double fading effect generated by introducing the ORIS module into the VLC channels.

Notation: Bold faces are used for vectors and matrices, $(\star)^T$ denotes the transpose operator, and $\mathbb{R}_+$ stands for the set of real-valued and non-negative numbers.

2 System Model

This study considers a multi-user indoor VLC system where ORIS modules enhance performance by optimizing light propagation. The system comprises light sources (LSs), photodetectors (PDs), and ORIS modules with multiple reflective elements. The focus is on analyzing the NLoS signal path facilitated by ORIS, ignoring less significant reflections from absorbing walls.

2.1 Indoor ORIS-Assisted VLC Model

We consider the multi-user VLC system illustrated in Fig. 1, where the ORIS infrastructure, with N elements, is implemented to improve the system's performance. L light sources (LSs) illuminate the environment within which K users connect to the network through the ORISs and LSs. In practice, three links prevail in this environment: (i) the LoS link, (ii) the NLoS path through the walls, and the NLoS channel through the ORIS infrastructure. In the sequel, we consider walls very absorbing, and the NLoS received signal through these does not add value to the overall system's performance. Therefore, we evaluate only the power transferred to the users through the NLoS paths with the ORIS module.

2.2 Channel Gains

In this paper, we model LSs as point sources, and the evaluated gains correspond to that of a point source. Note that in the remainder of this paper, NLoS means NLoS through the ORIS infrastructure.

The Line-of-Sight Path: The LoS gain, $h_{k,l}$, between the l -th LS and the k -th photodetector (PD) follows the Lambertian model. For a single light ray, $h_{k,l}$ can be derived as

$$h_{k,l} = \frac{A_{pd}(m+1)}{2\pi d_{k,l}^2} \cos^m(\theta_{l,k}) \cos(\phi_{l,k}) T(\phi_{l,k}) f(\phi_{l,k}). \quad (1)$$

In (1), A_{pd} , $d_{k,l}$, and m denote the PD effective area, the distance between the l -th LS and the k -th user, and the Lambertian index ($m = -1/\lceil \log_2(\cos(\beta_{1/2})) \rceil$ with $\beta_{1/2}$ as the half-power semi-angle of LSs), assuming that each receiver uses a single PD. $\theta_{l,k}$ and $\phi_{l,k}$ are the angle of irradiance and incidence, respectively. $f(\phi_{l,k})$ is the optical concentration gain, which considers the refractive index of the lens inside the PD, n_{pd} , and the receiver's field-of-view (FoV), Ψ_{pd} . $f(\phi_{l,k}) = n_{pd}^2 / \sin^2(\Psi_{pd})$ for $0 \leq \phi_{l,k} \leq \Psi_{pd}$ and $f(\phi_{l,k}) = 0$ otherwise. Finally, $T(\phi_{l,k})$ is the gain of the optical filter.

The Non-line-of-sight Paths: Considering the ORIS elements as point sources and that only first-order reflections are meaningful, the NLoS gain, $h_{k,n,l}$, between the k -user and the l -th LS through the n -th ORIS element can be expressed as

$$h_{k,n,l} = \frac{\partial_n A_n A_{pd}(m+1)}{2\pi (d_{l,n})^2 (d_{n,k})^2} \cos^m(\theta_{l,n}) \cos(\phi_{l,n}) \times \cos(\theta_{n,k}) \cos(\phi_{n,k}) T(\phi_{n,k}) f(\phi_{n,k}), \quad (2)$$

where ∂_n is the common reflectivity of the N ORIS elements, $d_{n,l}$ and $d_{k,n}$ are the distances between the l -th LS and the n -th ORIS element, and the n -th ORIS element and the k -th PD, respectively. A_n is the common effective area of the N ORIS elements, $\theta_{l,n}$ and $\theta_{n,k}$ are the angles of irradiance from the l -th LS towards the n -th ORIS element and from the n -th reflective surface towards the k -th user, and $\phi_{l,n}$ and $\phi_{n,k}$ are the angles of incidence on the n -th reflective surface and on the k -th user, respectively.

2.3 Received Signal Model and the Signal-to-Interference Plus Noise Ratio

Let $y_{k,l}$ be the signal the k -th PD receives when illuminated by the l -th LS. $y_{k,l}$ is defined by

$$y_{k,l} = y_{k,l} + y_{k,n,l} + z_k, \quad (3)$$

where $y_{k,l}$ and $y_{k,n,l}$ are the LoS and NLoS signals from the l -th LS to the k -PD, and z_k is a zero mean white additive Gaussian noise (AWGN) with variance σ_k^2 .

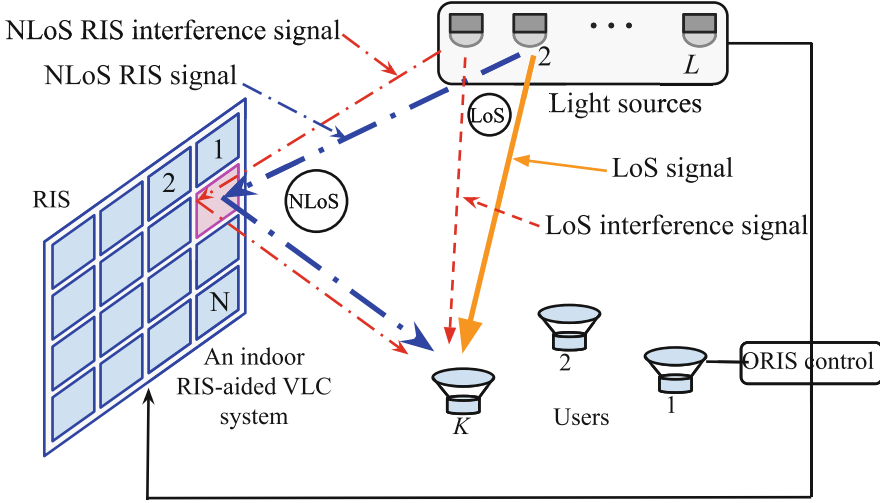


Fig. 1. A generalized ORIS-assisted indoor VLC system. L light sources, assisted by an ORIS module with N elements, illuminate a system of K users.

In indoor VLC systems, AWGN is a suitable model for both thermal and shot noises. The two signal components $y_{k,l}$ (LoS) and $y_{k,n,l}$ (NLoS) are respectively defined by

$$y_{k,l} = \rho_k h_{k,l} P_l x_l + \rho_l \sum_{i=1, i \neq l}^L h_{k,i} P_i x_i, \quad (4)$$

and

$$y_{k,n,l} = \rho_k \mathbf{h}_{k,n,l}^T \mathbf{g}_l P_l x_l. \quad (5)$$

In (5), $\mathbf{h}_{k,n,l} = [h_{k,1,l}, h_{k,2,l}, \dots, h_{k,N,l}] \in \mathbb{R}_+^{N \times 1}$ [4]. Equations (4) and (5) consider that the transmission of a symbol is defined by $\mathbf{x}^T = [x_1, x_2, \dots, x_L]$. ρ_k and P_l represent the PD's responsivity and the optical power the l -th LS emits, respectively. The quantity $\mathbf{g}_l$ belongs to the binary matrix $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_L]_{N \times L}$, where $\mathbf{G}$ describes the ORIS elements and LSs association [4]. The SINR between the l -th LS and the k -user can be represented as

$$\gamma_{k,l} = \frac{\rho_{pd}^2 \left[\frac{P_l}{q} \left(h_{k,l} + \mathbf{h}_{k,n,l}^T \mathbf{g}_l \right) \right]^2}{s_{k,l}}, \quad (6)$$

with $s_{k,l} = \sigma_k^2 + \rho_l^2 \sum_{i=1, i \neq l}^L h_{k,i}^2 P_i^2$. Note that an element of $\mathbf{g}_l$, $g_{l,n}$, is unity for an association of the l -th LS and the n -th reflective unit, and zero if not, and q and P_l are the ratios of the transmitted optical power to the electrical power, and the transmitted optical power, respectively.

Table 1. Simulation parameters.

Parameter	Value	Parameter	Value
B	200 MHz	$\beta_{1/2}$	70°
$\varsigma_{1,1}$	$10^{-21} \text{ A}^2/\text{Hz}$	ρ_{pd}	0.53 A/W
$f(\phi_{1,1})$	1	A_{pd}	1 cm^2
Ψ_{pd}	85°	g_{of}	1
P_l	1 W	λ	700 nm

3 Analysis and Numerical Results

3.1 Analysis

The capacity lower bound of the k -user in an ORIS-assisted indoor VLC data transmission system can be derived as

$$R_k = \sum_{l=1}^L \frac{1}{2} B \log_2 \left(1 + \frac{e}{2\pi} \gamma_{k,l} \right), \quad (7)$$

where e is the base of the natural logarithm and B is the signal-carrying bandwidth.

3.2 Comparison

The common reflectivity, ∂_n , of the N ORIS elements is the reflectivity at the boundary of the material used in the ORIS elements. It depends on the refractive indices, n_a , of the medium the light is traveling from, n_n , the refractive index of the medium the light is traveling into, the angle of incidence, $\theta_{l,n}$, and the angle of refraction, θ_n , of the signal through the reflective element. It is given by $\partial_n = 0.5\zeta_n$, where

$$\begin{aligned} \zeta_n = & \left(\frac{n_a \cos(\phi_{l,n}) - n_n \cos(\theta_n)}{n_a \cos(\phi_{l,n}) + n_n \cos(\theta_n)} \right)^2 \\ & + \left(\frac{n_a \cos(\theta_n) - n_n \cos(\phi_{l,n})}{n_a \cos(\theta_n) + n_n \cos(\phi_{l,n})} \right)^2. \end{aligned} \quad (8)$$

The first part of (8) relates to the s -polarized impinging light, while its second part is associated with a p -polarized incoming signal.

Figure 2 is a representation of both s - and p -polarized incident lights. $x-z$ is the plan of incidence. The s -polarized light is perpendicular to the plan of incidence, while the p -polarized light is parallel to the plan of incidence.

The above definition and description of the reflectivity of the reflecting material typically relate to sheet materials like glass mirrors and polished anodized aluminum, to mention only two. In the case of anisotropic materials such as LCs, the reflectivity, ∂_n , can be determined by considering the LC thickness that minimizes (or maximizes, depending on the situation) the low-field reflectance for a

Table 2. Comparison between potential materials to be used in optical ORISs for indoor VLC systems.

Material	Reflectivity	Control	Advantages	Disadvantages	Price (Internet)
Polished anodized Aluminum	$\approx 95\%$	based on micro-electro-mechanical systems (MEMS)	reflective, lightweight, structurally durable, and easily shaped	easily scratched, not super cheap	$\approx \$ 7$ per square foot
Mylar	$>95\%$	based on MEMS	super light, super cheap	requires a rigid backing	$\$ 30$ for a 4'x'50 roll
Aluminum foil	88%	based MEMS	extremely cheap and widely available	structurally weak	$\$3.29$ for 1.5'x70' roll
Can lids	75%	based on MEMS	widely Available, effective, salvageable	non-uniform shape, intensive labor to use	free
Acrylic mirror	99%	based on MEMS	very reflective, nearly unbreakable	super expensive	$\$78$ for 1'x4' plane
Glass mirror	99%	based on MEMS	super reflective, widespread-meaning it could be salvaged	very expensive and breaks easily	at least $\$100$ /square foot
Astro-foil	76%	based on MEMS	relatively strong	not so reflective	$\$37$
Liquid crystals and other phase-changing materials (refractive index, n_{lc})	n_{lc} and external voltage related	based on externally applied field and refractive index-MEMS	high flexibility in beam management	may reduce transmission speed	vary with the LC' substance

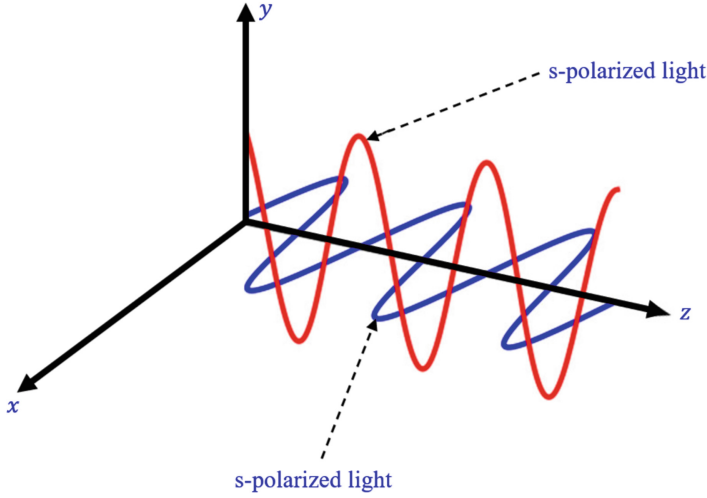


Fig. 2. Diagram showing *s*-polarized and *p*-polarized light of the incident light.

given wavelength and incidence angle, $\theta_{l,n}$. In this case, we evaluate the reflectivity of the ORIS elements as

$$\partial_n = \frac{\zeta_n \sin^2 \left(\frac{2\pi}{\lambda} n_{lc} a_{lc} \psi(\theta_{l,n}) \right)}{2 + \zeta_n \sin^2 \left(\frac{2\pi}{\lambda} n_{lc} a_{lc} \psi(\theta_{l,n}) \right)}, \quad (9)$$

where λ , a_{lc} , and $\theta_{l,n}$ are the incident light wavelength, the LC's thickness, and the light angle after the first surface transmission.

3.3 Analytical Results

Table 1, where ρ_{pd} is the responsivity of the PD, depicts the system parameters used in the simulation, while Table 2 shows the reflecting or transition coefficient of several materials that can be considered in the design of optical ORISs for indoor VLC systems.

We analyze parameters such as transition coefficient, ∂_n , maneuverability, advantages, disadvantages, and price collected from the internet. We consider the following system's parameters to show the performance of ORIS-assisted indoor VLC systems based on these materials. We consider a $5\text{ m} \times 5\text{ m} \times 3\text{ m}$ room model with a single LS, $L = 1$, one user, $K = 1$, and an ORIS with 100 reflective elements of $0.1\text{ m} \times 0.1\text{ m}$ each.

The NLoS ORIS channel gain $h_{l,n,k}$ in (2) can be reduced to $h_{1,n,1} = \partial_n \Delta$ for one LS, 100 ORIS elements, and one user. With the numerical parameters of the system, assuming a central position of the PD facing the ceiling, the LS fixed on the top line of the back wall, and the ORIS module located at the center of the wall on the left-hand side of the room, we obtain $d_{1,n} = 3.84\text{ m}$, $d_{n,1} = 2.92\text{ m}$, $\cos(\theta_{1,n}) = 0.39$, $\cos(\phi_{1,n}) = 0.92$, $\cos(\theta_{n,1}) = 0.514$, and $\cos(\phi_{n,1}) = 0.92$, then

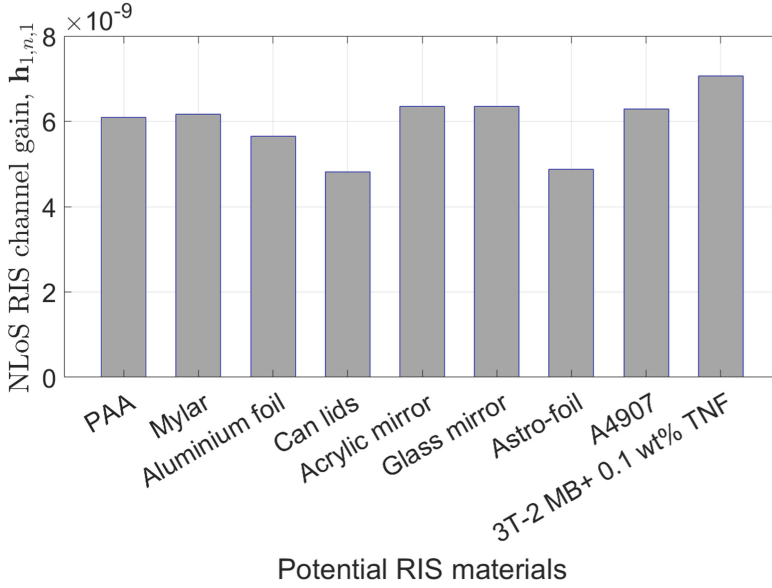


Fig. 3. NLoS ORIS channel gain related to several selected potential ORIS materials.

we have $\Delta = 6.419 \times 10^{-11}$. With the same parameters and the knowledge that the air refractive index is unity, $n_a = 1.000$, and exploiting (8), we realize $\partial_n = 0.95, 0.88, 0.75, 0.99, 0.98, 0.76$ for polished anodized Aluminium, Aluminum foil, can lids, acrylic mirror, glass mirror, and Astro foil, respectively. In addition, we obtain the reflectivity of Mylar, which is higher than 0.95. Furthermore, we evaluate that of substances like A4907 and 3T-2MB combined with 0.01 wt% of TNF LCs using (9), and obtain several values.

Figures 3 and 4 depict the channel gain and the achievable rate that relates to these potential materials, which could be used to design ORIS modules for the indoor environment. For the same transmission parameters, environment, reflecting surface orientation, and receiver orientation, the channel gain varies with the material used, as this depends on the reflectivity of this material. Figure 3 shows that among the sheet material investigated in this paper, mirrors have the highest reflectivity, around 0.99, which means that only 1% of the impinging signal is lost during the transition through the ORIS reflective surface. Up to 25%

Excepting the reflectivity, Table 2 also shows more critical parameters in selecting the ORIS material. These are the maneuverability/control, main advantages and disadvantages, and cost. Generally, in ORIS modules, The control mechanism may use micro-electro-mechanical systems (MEMS) or ultra-fast switching elements that interact with the ORIS controller (see Fig. 1).

Due to their solid state, most materials listed in Table 2 do not allow their intrinsic characteristics to change during their operations. Therefore, their tuning solely uses the MEMS technique. Orienting the impinging signal toward a

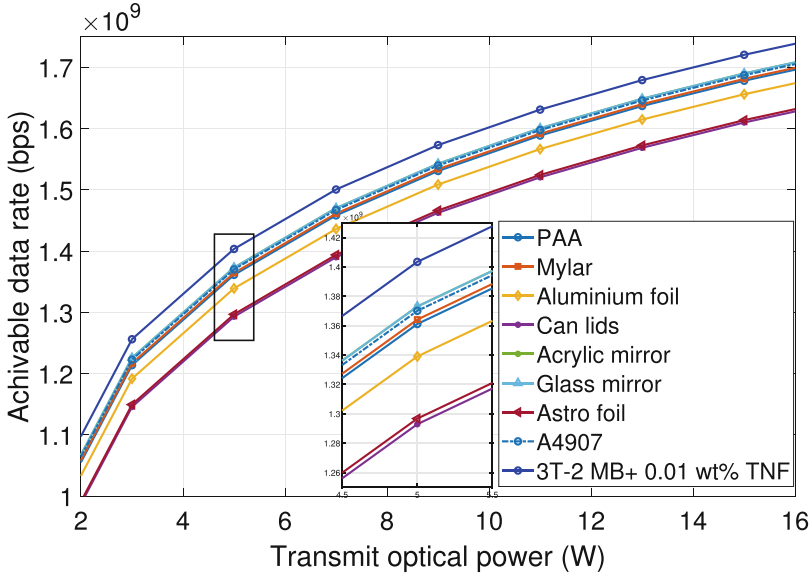


Fig. 4. Achievable data rate vs. transmit optical power of IRS-aided indoor VLC systems: Comparison between several potential ORIS materials.

specific user requires a rotation of the element (yaw and roll angles). Some materials, like PAA, can be readily shaped, scratched easily, and expensive. Mylar is very cheap, about \$30 for a 4'x'50 roll; however, it requires extra material for the framing, which will increase its cost. Distinctively, Aluminum foil is immensely cheap and widely accessible but structurally weak. Although can lids are effective, widely available for free, and salvageable, they may be presented in several different shapes and require extra labor to be usable. While Acrylic mirrors are very reflective and expensive, glass mirrors are also highly reflective, costly but breakable. On the other hand, ORIS elements using anisotropic substances like LCs do not require a unit rotation to re-direct the light rays. In this case, the orientation of the incoming light is based on Snell's law of reflection, depending on tuning the refractive index using an externally applied field.

Of all the materials listed in Table 2, LCs can be controlled with MEMS and externally applied electric or magnetic fields. They present high flexibility in beam management; however, due to the tuning time related to the alignment of their directors, LCs may slow down transmission. To better contextualize the results, we propose elaborating on how the variations in material reflectivity impact real-world ORIS module designs for VLC systems. For instance, materials such as polished anodized aluminum exhibit high reflectivity (95%) and are ideal for high-capacity indoor networks where signal fidelity is paramount. However, recycled materials like can lids, while less reflective (75%), offer an environmentally friendly and cost-effective solution for non-critical or tempo-

rary installations. This practical perspective bridges the gap between theoretical findings and their tangible applications.

4 Conclusion

This paper has investigated several potential ORIS materials for indoor VLC systems. The reflectivity of these materials has been determined, and the corresponding ORIS-assisted indoor VLC channel gains have been evaluated. While using the investigated materials, the achievable rates of ORIS-assisted indoor VLC systems with single LS, ORIS with 100 elements, and single user are provided, showing the effect of these materials on the data transmission. Materials with low reflectivity lead to low transmission rates and vice-versa. LCs present the advantage of emerging light manipulation by controlling the refractive index without module rotation. Their high ability to electrically tune the emerging light beam and their amplification capability are two of their most critical advantages that could be traded off with slow data transmission. The findings of this study underscore the critical role of material selection in ORIS-assisted VLC systems. High-reflectivity materials like polished anodized aluminum and acrylic mirrors are suitable for applications requiring maximum channel gain and achievable rates. On the other hand, recycled materials present an eco-friendly alternative for budget-conscious deployments. Liquid crystal-based ORIS stands out as a versatile solution, enabling superior reflectivity control and beam management, making it a promising candidate for future dynamic VLC networks. These insights provide a roadmap for designing ORIS modules tailored to specific real-world scenarios.

References

1. Ndjiongue, A.R., Ferreira, H.C., Ngatched, T.M.: Visible light communications (VLC) technology. In: Wiley Encyclopedia of Electrical and Electronics Engineering, pp. 1–5 (2015)
2. Qian, S.: Visible light communications with image sensors. In: Proceedings of 2024 9th Optoelectronics Global Conference (OGC), pp. 69–72 (2024)
3. Guzman, B.G., Mir, M.S., Fonseca, D.F., Galisteo, A., Wang, Q., Giustiniano, D.: Prototyping visible light communication for the internet of things using Openvlc. *IEEE Commun. Mag.* **61**(5), 122–128 (2023)
4. Sun, S., Yang, F., Song, J., Han, Z.: Optimization on multiuser physical layer security of intelligent reflecting surface-aided VLC. *IEEE Wireless Commun. Lett.* **11**(7), 1344–1348 (2022)
5. Aboagye, S., Ngatched, T., Dobre, O.A., Ndjiongue, A.R.: Intelligent reflecting surface-aided indoor visible light communication systems. *IEEE Commun. Lett.* **25**(12), 3913–3917 (2021)
6. Ndjiongue, A.R., Ngatched, T., Dobre, O.A., Haas, H.: Digital RIS (DRIS): the future of digital beam management in RIS-assisted OWC systems. *J. Lightwave Technol.* **40**(16), 5597–5604 (2022)

7. Abumarshoud, H., Mohjazi, L., Dobre, O.A., Renzo, M.D., Imran, M.A., Haas, H.: LiFi through reconfigurable intelligent surfaces: a new frontier for 6G? *IEEE Veh. Technol. Mag.* **17**(1), 37–46 (2021)
8. Cang, Y., et al.: Joint deployment and resource management for VLC-enabled RISs-assisted UAV networks. *IEEE Trans. Wireless Commun.* **22**(2), 746–760 (2022)
9. Liu, C., Yu, L., Yu, X., Qian, J., Wang, Y., Wang, Z.: Capacity analysis of RIS-assisted visible light communication systems with hybrid NOMA. In: *Proceedings of IEEE Globecom Workshops (GC Wkshps)*, pp. 118–123 (2022)
10. Sun, S., Yang, F., Song, J., Zhang, R.: Intelligent reflecting surface for MIMO VLC: joint design of surface configuration and transceiver signal processing. *IEEE Trans. Wireless Commun.* **22**(9), 5785–5799 (2023)
11. Abdelhady, A., Amin, O., Salem, A., Alouini, M.-S., Shihada, B.: Channel characterization of IRS-based visible light communication systems. *IEEE Trans. Commun.* **70**(3), 1913–1926 (2022)
12. Abdelhady, A.M., Salem, A., Amin, O., Shihada, B., Alouini, M.-S.: Visible light communications via intelligent reflecting surfaces: metasurfaces vs mirror arrays. *IEEE Open J. Commun. Soc.* **2**, 1–20 (2020)
13. ElMossallamy, M.A., Zhang, H., Song, L., Seddik, K.G., Han, Z., Li, G.Y.: Reconfigurable intelligent surfaces for wireless communications: principles, challenges, and opportunities. *IEEE Trans. Cogn. Commun. Netw.* **6**(3), 990–1002 (2020)
14. Basar, E.: Reconfigurable intelligent surface-based index modulation: a new beyond MIMO paradigm for 6G. *IEEE Trans. Commun.* **68**(5), 3187–3196 (2020)
15. Ndjiongue, A.R., Ngatched, T., Dobre, O.A., Haas, H.: Toward the use of reconfigurable intelligent surfaces in VLC systems: beam steering. *IEEE Wireless Commun.* **28**(3), 156–162 (2021)
16. Ndjiongue, A.R., Ngatched, T., Dobre, O.A.: Impact of the refractive index on the achievable rate of liquid crystal-based digital-RIS indoor VLC systems. *IEEE Photon. J.* **15**(1), 1–6 (2022)



Single- vs Multi-connectivity in Movable Access Point-Enabled Visible Light Communication Systems

Zohaib Saghir¹(✉), Raphael Ahiaklo-Kuz², and Sylvester Aboagye¹ 

¹ School of Engineering, University of Guelph, Guelph, ON N1C 1A7, Canada
zsaghir@uoguelph.ca, saboagye@uoguelph.ca

² Department of Electrical and Computer Engineering, McMaster University,
Hamilton, ON L8S 4L8, Canada
ahiaklor@mcmaster.ca

Abstract. Visible light communication (VLC) is among the key enabling technologies to shape future wireless networks due to the unlicensed and vast amount of bandwidth in the visible light spectrum. Unlike radio frequency (RF) systems, signal transmission in VLC is severely affected by line-of-sight (LoS) link blockages, random device orientation, and limited field-of-view (FoV) of VLC transceivers, which lead to non-uniform coverage, reduced data rate and limited mobility support. Recently, the idea of movable access points (MAPs) in VLC systems has been proposed to overcome these challenges. This paper investigates the concepts of single- and multi-connectivity in MAP-based VLC systems. Specifically, two sum rate optimization frameworks for single- and multi-connectivity-enabled MAP-based VLC system are proposed, while considering minimum quality-of-service requirements for communication and sensing users and power budget. Then, the majorization-minimization approach is leveraged to develop an algorithm for the frameworks. Finally, simulation results are used to evaluate the sum rate and computational complexity of the two frameworks. The results reveal that single- and multi-connectivity frameworks mitigate the impact of LoS blockages, user mobility, and limited FoV of transceivers in VLC systems and a trade-off between complexity and sum rate, where multi-connectivity offers a 300–400% improvement, but a 2.4x increase in computational complexity compared to single-connectivity.

Keywords: Visible light communication · movable access point · illumination · user mobility · sum rate optimization

1 Introduction

Visible light communication (VLC) is a promising technology to complement existing radio frequency (RF)-based communication systems (e.g., wireless

This work was supported by University of Guelph.

fidelity, sub-6 GHz, and millimeter wave) in sixth generation (6G) and beyond networks due to its vast and unlicensed spectrum and low-cost infrastructure [1]. However, VLC is highly susceptible to line-of-sight (LoS) link blockage, which limits its coverage and data rate performance. In addition, the field-of-view (FoV), random device orientation, and the limited illumination coverage of access points (APs) also affects mobility support and the performance of VLC systems, in general.

Optical reconfigurable intelligent surfaces (ORISs) have been proposed in the literature to tackle the issues of LoS link blockages, limited FoV, and random device orientation [1–8]. ORISs mitigate the impact of these design issues by reflecting incident light beams from the AP toward the receiver. Thus, it establishes an ORIS-assisted non-line-of-sight (NLoS) path between APs and receivers. However, it has been established that the link performance (e.g., data rate) of the ORIS NLoS links is insignificant when compared to that of the direct LoS link between the AP and the receiver.

Recently, the idea of movable access point (MAP) technology in indoor VLC systems was proposed in [9] to tackle LoS link blockages, the limited FoV of transceivers, and random device orientation while providing mobility support. Similar to motorized track lighting fixtures that use direct current motors or actuators, the light fixtures in a MAP-based VLC system use adaptors and mounting arms to move along a track in two-dimensional (2D) directions to optimize coverage, support user mobility, and enhance network performance. In that paper, the authors studied the joint optimization of MAPs positioning and users association to maximize the aggregate communication sum rate and sensing mutual information (MI) for an integrated visible light communication and sensing (VLCS) system.

Unlike OIRS which has been widely investigated, several issues remain to be studied for indoor MAP-based visible light networks. First, since ceiling-mounted APs can be moved in a MAP-based system, users can be potentially served by single or multiple APs simultaneously via single- and multi-connectivity, respectively, depending on the distribution of blockers and link blockages, and users' quality-of-service (QoS) requirements. Multi-connectivity can be used to enhance data rate and MI to support applications with high QoS requirements such as ultra high definition video and indoor precise positioning and navigation [10]. Thus, multi-connectivity can help increase network availability, resiliency, and reliability. However, there has been no study on the design and optimization of single- and multi-connectivity based MAP-based communication-centric VLCS systems. Second, in [9], illumination constraints necessary for compliance with eye safety standards for visible light networks were not considered [11].

Motivated by the above, this paper investigates one method to make communication and sensing more efficient by exploiting single- and multi-connectivity in MAP-based communication-centric VLCS systems. A novel optimization problem is formulated under single- and multi-connectivity constraints, and a low-complexity solution framework is developed to obtain a locally optimal solution. The main contributions are summarized as follows:

- We consider a MAP-based communication-centric VLCS system and propose two optimization frameworks, namely, single- and multi-connectivity frameworks that optimize the position of MAPs and user-AP(s) association to maximize the total sum rate.
- We implement MAPs positioning and a user-AP association strategy based on the majorization-minimization (MM) method to showcase the effectiveness of single- and multi-connectivity optimization frameworks in overcoming link blockages, limited FoV, random device orientation, and supporting user mobility.
- We examine the impact that system parameters such as user mobility and number of users, blockers, and grid-points have on the sum rate.

2 System Model

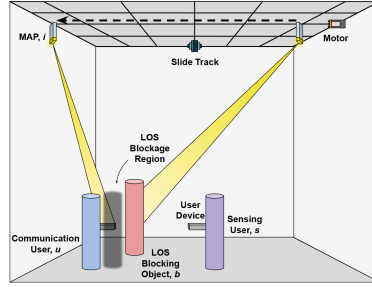


Fig. 1. MAP-aided VLCS system with communication and sensing users and LoS blocking objects.

2.1 Network Model

The system model of the MAP-based VLCS system is illustrated in Fig. 1, with a focus on downlink transmission. The set of MAPs is denoted as $\mathcal{I} = \{1, 2, \dots, I\}$. Multiple mobile communication and sensing users, and blockers are uniformly and randomly distributed in the indoor environment, and represented as $\mathcal{U} = \{1, 2, \dots, U\}$, $\mathcal{S} = \{1, 2, \dots, S\}$, and $\mathcal{B} = \{1, 2, \dots, B\}$, respectively, with $\mathcal{K} = \{\mathcal{S} \cup \mathcal{U}\}$, where $k \in \mathcal{K}$, and $\mathcal{S} \cap \mathcal{U} = \emptyset$. The room ceiling is discretized into grid points, denoted as $\mathcal{G} = \{1, 2, \dots, G\}$, for the movement of the MAPs. Similar to [9], each MAP is equipped with a light emitting diode (LED) and a photodetector (PD) for simultaneous signal transmission and active sensing. Each receiver has a PD for signal reception. User mobility is in accordance with the random way point mobility model [12].

2.2 Channel Model

In VLC, only the LoS paths are considered since the NLoS paths from diffused reflections have weak signal power. The channel gain from MAP i to user k when there is no link blockage is expressed as [5, 9]:

$$g_{i_k,g} = \begin{cases} \frac{(\eta+1)A}{2\pi D_{i_k,g}^2} \cos^\eta(\phi_{i_k,g}) \mathcal{RT}(\theta_{i_k,g}) \cos(\theta_{i_k,g}), & 0 \leq \theta_{i_k,g} \leq \mathcal{F} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where an object between a user and a MAP does not block the LoS path if the height of the object $h_b < \mathcal{H}$ as shown in Fig. 2, where

$$\mathcal{H} = \frac{x_{b,j}h_r + (d_x - x_i h_{MAP})}{d_x}, \quad (2)$$

is the height threshold, h_{MAP} and h_r are the heights of the MAP and receiver respectively, $x_{b,j}$ denotes the horizontal distance between the MAP and the blocking object (user or non-user) and d_x is the separation distance between MAP and the user device.

In (1), A is the physical surface area of the PD, $\phi_{i_k,g}$ is the angle of irradiance of MAP i located at grid point g on user k , $\theta_{i_k,g}$ is the angle of incidence of the PD, $\mathcal{R}$ is the optical filter gain and $\mathcal{T}(\theta_{i_k,g}) = (f^2/\sin^2 \mathcal{F})$ is the optical concentrator gain, where f is the refractive index, $D_{i_k,g}$ is the distance from MAP i to user k and $\mathcal{F}$ is the FoV of the MAP transceiver. $\eta = -1/\log_2(\cos \Theta_{1/2})$ is the Lambertian index, with $\Theta_{1/2}$ being the half-intensity radiation angle. Device orientation of the user device is modeled in terms of the device's polar angle, α , and azimuth angle, β , as [13]:

$$\cos(\theta_{i_k,g}) = \left(\frac{x_i - x_{i_k,g}}{D_{i_k,g}} \right) \sin(\Gamma) \cos(\Omega) + \left(\frac{y_i - y_{i_k,g}}{D_{i_k,g}} \right) \sin(\Gamma) \sin(\Omega) + \left(\frac{z_i - z_{i_k,g}}{D_{i_k,g}} \right) \cos(\Omega), \quad (3)$$

where (x_i, y_i, z_i) and $(x_{i_k,g}, y_{i_k,g}, z_{i_k,g})$ are position vectors describing the MAP and user's locations, respectively, where Γ can be modeled using the truncated Laplace distribution with the mean and standard deviation of 41° and 9° , respectively. Ω follows a uniform distribution with $\Omega \sim U[-\pi, \pi]$ [13].

2.3 Illumination in VLC

The luminous flux of the unit optical power ϕ_v is defined as [14]:

$$\phi_v = \frac{(\eta+1)A}{2\pi D_{i_k,g}^2} \delta \cos^\eta(\phi_{i_k,g}) \cos(\theta_{i_k,g}), \quad (4)$$

where δ is the optical power to luminous flux conversion factor. The illumination at any point can be expressed as $E = \phi_v \zeta \sqrt{\mathcal{P}_T}$, where ζ is a conversion factor and $\mathcal{P}_T$ is the electrical transmit power.

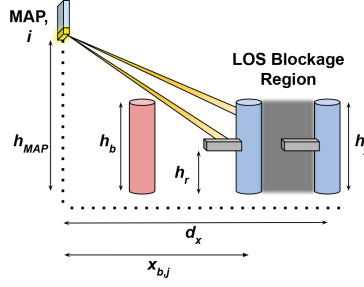


Fig. 2. LoS blockage model.

3 Sum Rate Optimization

Two performance metrics are considered for the communication-centric VLCS, namely, data rate for communication users and mutual information for sensing targets. The signal-to-noise ratio is the ratio between the received visible light signal power signal and receiver noise in the indoor environment and is defined as $\mathcal{P}_T(R_{PD} g_{i_k,g})^2 / \mathcal{W}N_{VLC}$, where $\mathcal{W}$ is the bandwidth, R_{PD} is the responsivity of the PD, and N_{VLC} is the power spectral density of additive white Gaussian noise. Therefore, the achievable data rate received by communication user u from the MAP i at grid g can be given by [5]:

$$R_{i_u,g} = \mathcal{W} \log_2 \left(1 + \frac{e}{2\pi} \frac{\mathcal{P}_T(R_{PD} g_{i_k,g})^2}{\mathcal{W}N_{VLC}} \right), \quad (5)$$

Furthermore, the received echo signal-to-noise ratio at the ceiling for sensing user s is expressed as $(e/2\pi) (\mathcal{P}_T \Lambda g_{i_s,g}^2 / \mathcal{W}N_{VLC})$, where Λ is the reflecting loss factor, $g_{i_s,g}^2$ is the double path channel gain for the links from MAP i to user s and user to the same MAP. The MI at the MAP i located at grid g for sensing target s can be presented as [15]:

$$R_{i_s,g} = \mathcal{W} \log_2 \left(1 + \frac{e}{2\pi} \frac{\mathcal{P}_T \Lambda g_{i_s,g}^2}{\mathcal{W}N_{VLC}} \right). \quad (6)$$

3.1 Single-Connectivity Optimization Framework

The single-connectivity optimization framework to maximize the sum data rate by jointly optimizing MAP positioning and user-AP association under minimum QoS requirements for sensing and communication users and transmit power budget can be formulated as

$$\begin{aligned}
& \max_{\mathbf{v}} \sum_{g=1}^G \sum_{i=1}^I \sum_{u=1}^U v_{i_u,g} \mathcal{W} \log_2 \left(1 + \frac{e}{2\pi} \frac{\mathcal{P}_T(R_{\text{PD}} g_{i_k,g})^2}{\mathcal{W} N_{\text{VLC}}} \right) \\
& \text{s.t.} \\
& \mathbf{C}_1 : \sum_{g=1}^G \sum_{i=1}^I v_{i_u,g} \mathcal{W} \log_2 \left(1 + \frac{e}{2\pi} \frac{\mathcal{P}_T(R_{\text{PD}} g_{i_k,g})^2}{\mathcal{W} N_{\text{VLC}}} \right) \geq R_c, \forall u, \\
& \mathbf{C}_2 : \sum_{g=1}^G \sum_{i=1}^I v_{i_s,g} \mathcal{W} \log_2 \left(1 + \frac{e}{2\pi} \frac{\mathcal{P}_T \Lambda g_{i_s,g}^2}{\mathcal{W} N_{\text{VLC}}} \right) \geq R_t, \forall s, \\
& \mathbf{C}_3 : \sum_{g=1}^G \left(\sum_{u=1}^U v_{i_u,g} + \sum_{s=1}^S v_{i_s,g} \right) \leq 1, \forall i, \\
& \mathbf{C}_4 : \sum_{i=1}^I \left(\sum_{u=1}^U v_{i_u,g} + \sum_{s=1}^S v_{i_s,g} \right) \leq 1, \forall g, \\
& \mathbf{C}_5 : \sum_{g=1}^G \sum_{i=1}^I v_{i_u,g} = 1, \forall u \\
& \mathbf{C}_6 : \sum_{g=1}^G \sum_{i=1}^I v_{i_s,g} = 1, \forall s \\
& \mathbf{C}_7 : \left(\frac{(m+1)}{2\pi D_{i_k,g} \delta} \cos(\phi_{i_k,g})^m \cos(\theta_{i_k,g}) \zeta \right)^2 \mathcal{P}_T \\
& \quad \leq (E_{\max})^2 v_{i_k,g}, \forall j \\
& \mathbf{C}_8 : \left(\frac{(m+1)}{2\pi D_{i_k,g} \delta} \cos(\phi_{i_k,g})^m \cos(\theta_{i_k,g}) \zeta \right)^2 \mathcal{P}_T \\
& \quad \geq (E_{\min})^2 v_{i_k,g}, \forall j \\
& \mathbf{C}_9 : v_{i_u,g} \in \{0, 1\}, a_{i_s,g} \in \{0, 1\}, \forall i, g, u, s.
\end{aligned} \tag{7}$$

where $v_{i_k,g}$ is a binary variable that indicates whether user k is being served by MAP i located at grid point g or not. The constraints in the problem (7) are explained as follows. Constraints $\mathbf{C}_1$ and $\mathbf{C}_2$ ensure that QoS requirements for communication and sensing users are satisfied. Constraint $\mathbf{C}_3$ ensures that MAP i can be placed at only one single grid point. Constraint $\mathbf{C}_4$ ensures that only one MAP can be positioned at grid point g . Constraints $\mathbf{C}_5$ and $\mathbf{C}_6$ guarantee single-connectivity between users and MAPs. Constraints $\mathbf{C}_7$ and $\mathbf{C}_8$ denote the maximum and minimum illumination requirements for all users. Finally, $\mathbf{C}_9$ ensures that the decision variables for MAP positioning and user association are binary.

3.2 Multi-connectivity Optimization

The following modifications can be made to the problem in (7) to obtain an equivalent optimization framework under multi-connectivity constraints. Specifically $\mathbf{C}_5$, and $\mathbf{C}_6$ of (7) can be written as

$$\begin{aligned}
\mathbf{C}_5 : 1 & \leq \sum_{g=1}^G \sum_{i=1}^I a_{i_u,g} \leq N_{\text{MAP}}, \forall u \\
\mathbf{C}_6 : 1 & \leq \sum_{g=1}^G \sum_{i=1}^I a_{i_s,g} \leq N_{\text{MAP}}, \forall s,
\end{aligned} \tag{8}$$

Algorithm 1 Algorithm for solving (10) [9]

-
- 1: **Initialization:** maximum number of iterations $\kappa_{\max}$, $\Psi \gg 1$, $\kappa = 1$, and $\mathbf{v}^{(0)}$;
 - 2: **Repeat**
 - 3: Solve (10) for a given $\mathbf{v}^{(\kappa-1)}$ and obtain $\mathbf{v}^\kappa$;
 - 4: Set $\kappa = \kappa + 1$;
 - 5: **Until** convergence or $\kappa = \kappa_{\max}$;
-

where constraints $\mathbf{C}_5$ and $\mathbf{C}_6$ are modified to allow a user to connect to multiple MAPs simultaneously, where N_{MAP} is the total number of MAPs.

3.3 Proposed Solution

The optimization problems in (7) and (8) fall into the category of combinatorial optimization which has been proven to be a non-deterministic polynomial (NP)-hard problem. In order to demonstrate the performance of the two optimization frameworks, we employ a MM-based solution approach to obtain a locally optimal solution [9]. According to this approach, the binary constraints in $\mathbf{C}_9$ can be relaxed into the equivalent continuous form:

$$\begin{aligned} \mathbf{C}_{9a} : 0 &\leq a_{i_k,g} \leq 1, \\ \mathbf{C}_{9b} : \sum_{i=1}^I \sum_{k=1}^K \sum_{g=1}^G a_{i_k,g} - \sum_{i=1}^I \sum_{k=1}^K \sum_{g=1}^G a_{i_k,g}^2 &\leq 0, \end{aligned} \quad (9)$$

and $\mathbf{C}_9$ can be replaced by $\mathbf{C}_{9a}$ and $\mathbf{C}_{9b}$. Then, the objective function can be in (7) can be replaced by the approximated convex function

$$\max_{\mathbf{V}} \sum_{g=1}^G \sum_{i=1}^I \left(\sum_{u=1}^U v_{i_u,g} \mathcal{W} \log_2 \left(1 + \frac{e}{2\pi} \frac{\mathcal{P}_T(R_{\text{PD}} g_{i_k,g})^2}{\mathcal{W} N_{\text{VLC}}} \right) \right) - \Psi \left(\sum_{i=1}^I \sum_{k=1}^K \sum_{g=1}^G \left(v_{i_k,g} (1 - 2v_{i_k,g}^{(\kappa-1)}) \right) \right), \quad (10)$$

where $v_{i_k,g}^{(\kappa-1)}$ is the $(\kappa - 1)$ -th iterate solution of the problem, $\Psi \gg 0$ is a large penalty factor, and $\left(\sum_{i=1}^I \sum_{k=1}^K \sum_{g=1}^G \left(v_{i_k,g} (1 - 2v_{i_k,g}^{(\kappa-1)}) \right) \right)$ is the approximated penalty function after leveraging the MM technique [16]. The communication-centric optimization problem with objective function (10) and constraints $\mathbf{C}_1 \sim \mathbf{C}_{9a}$ is convex and can be iteratively solved to obtain a locally optimal solution using Algorithm 1.

4 Simulation Results

The indoor environment is a 10 m $\times$ 10 m $\times$ 3 m model that consists of multiple users and non-user blockers with cylindrical dimensions of a 0.15 m radius, 1.65 m height, and a user-to-device distance of 0.36 m. A traditional VLC system where an AP is fixed on the ceiling at the room center, named as ‘‘Benchmark’’ scheme is considered for comparison purposes. The number of LED arrays for

Benchmark is the same as the combined number of LED arrays for the MAPs. For example, if the proposed scheme has 20 APs with $\mathcal{P}_T = 1$ W dispersed at varying grid-point locations, the Benchmark scheme would be designed as a single AP source with transmit power $\mathcal{P}_T = 20$ W. MATLAB R2024b was used in accordance with the simulation parameters defined in Table 1 to obtain the simulation results.

Table 1. Simulation Parameters

Parameter	Value	Parameter	Value
User's Speed	0.5–2 m/s	R_c	5–10 Gb/s
R_t	1–4 Gb/s	$\kappa_{\max}$	30
Time Slots per Instance	10	G	100
I	20	B	4
A	1 cm ²	Υ	1
$\mathcal{F}$	70°	$\Theta_{1/2}$	60°
f	1.5	$\mathcal{W}$	5 GHz
$\mathcal{P}_T$	1 W	R_{PD}	0.53 A/W
N_{VLC}	10^{-21} A ² /Hz Ψ		10^6
v_{max}	20	ζ	3
δ	2.1×10^{-3}	E_{min}	10
E_{max}	800		

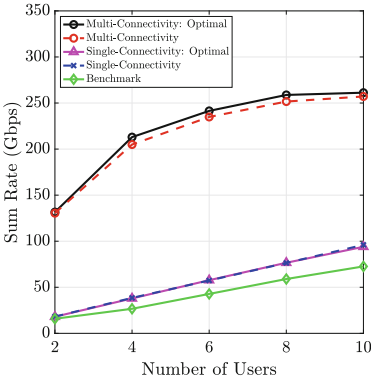


Fig. 3. Sum rate vs. number of users.

Figure 3 shows the sum rate performance for multi-connectivity and single-connectivity optimization frameworks as well as the Benchmark scheme under different number of users. It can be observed that the multi-connectivity framework significantly outperforms the single-connectivity scheme by a gain of about

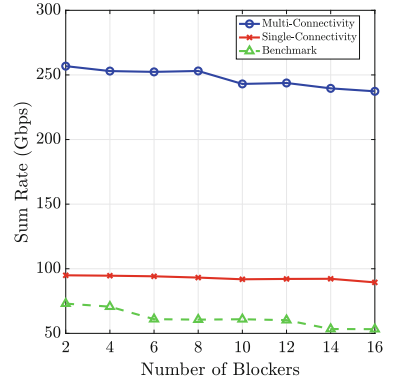


Fig. 4. Sum rate vs. number of blockers.

365.48% due to increased AP(s)-user connectivity. It is revealed that the rate of increase in sum rate as the number of users grows is much steeper for the multi-connectivity framework and then begins to saturate when the number of users is 8. However, the rate of increase for single-connectivity framework is linear with increasing number of users. Both single- and multi-connectivity achieve superior sum rate when compared to the Benchmark scheme. Moreover, the solution of Algorithm 1 is very close (i.e., near optimal) to the globally optimal solution for both frameworks. Specifically, performance gaps of about 1% and less than 3% are noticed for single- and multi-connectivity, respectively. The difference in the performance gap is because of the larger optimization problem dimension for the multi-connectivity framework when compared to the single-connectivity framework.

Figure 4 depicts the sum rate performance for single- and multi-connectivity frameworks and the Benchmark scheme for varying number of blockers when there are 10 users (5 communication and 5 sensing users). While the sum rate of single- and multi-connectivity optimization frameworks decreases slowly with increasing number of blockers, that of the Benchmark decreases rapidly. This observation motivates MAP technology. Moreover, multi-connectivity significantly outperforms single-connectivity.

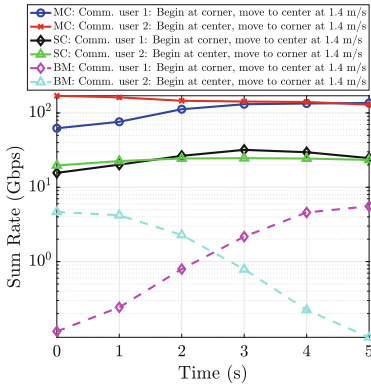


Fig. 5. Impact of user mobility on sum rate.

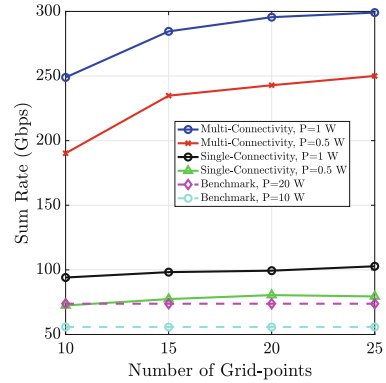


Fig. 6. Sum rate vs. number of grid-points.

The influence of user mobility on sum rate is displayed in Fig. 5 for single- and multi-connectivity schemes and the Benchmark. For this figure, we use “MC”, “SC”, and “BM” for multi-connectivity, single-connectivity, and Benchmark, respectively. The Benchmark scheme exhibits notable data rate fluctuations due to its fixed AP position and coverage. By contrast, the MAP-based MC and SC configurations adjusted dynamically to user movement, consistently optimizing APs placements to maintain coverage and high data rate across various users’ locations and achieving stable data rates.

In Fig. 6, the effect of grid-point density on sum rate is analyzed for single- and multi-connectivity schemes at various transmit power levels. It can be

observed that augmenting the transmit power leads to significant improvements in sum rate for all frameworks. This is due to the increase in the feasible region of the optimization problem. In the MAP-based systems, denser grid points initially contributed to an enhanced sum rate by providing more options for placing the APs. However, the rate of increase reduces with increasing number of grid points, stabilizing around $G = 25$ grid points due to diminishing returns. The Benchmark system, unaffected by grid density changes, showed no significant variation in performance due to its fixed AP configuration.

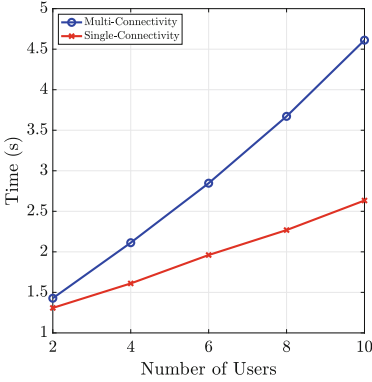


Fig. 7. Time complexity for Algorithm 1.

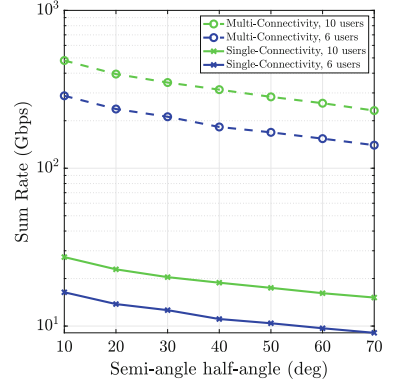


Fig. 8. Sum rate vs. LED half-angle.

Figure 7 illustrates the time complexity of implementing Algorithm 1 for single- and multi-connectivity optimization frameworks. It can be observed that the complexity of both frameworks increases almost linearly with increasing number of users. The multi-connectivity framework has higher complexity than single-connectivity due to the larger number of variables and constraints involved. The gap between the complexity for single- and multi-connectivity increases as the number of users get higher, with multi-connectivity systems exhibiting a 2.4x increase in computational complexity. In order to reduce the computational complexity, recent machine learning algorithms can be exploited to develop an algorithm with a significantly reduced complexity.

Finally, Fig. 8 provides a plot of sum rate at various half-angles for single- and multi-connectivity with 6 and 10 user scenarios. It demonstrates that multi-connectivity consistently achieves a higher sum rate across all half-angle values due to the simultaneous utilization of multiple APs for each user. As observed, the sum rate in both connectivity schemes decreases as the half-angle increases from 10° to 70° . This decrease is primarily due to the light beam dispersion at larger half-angles, which reduces the effective channel gain per AP-user link, thus impacting the overall sum rate. For multi-connectivity, the aggregation of signal contributions from multiple APs partially mitigates this effect, maintaining a higher sum rate across the range of half-angles. In comparison, single-connectivity, where each user connects to only one AP, shows a more pronounced decline in sum rate with increasing half-angle. This limitation is attributed to the

inability of single-connectivity to leverage additional APs to offset the reduced gain at wider angles. On average, multi-connectivity achieves approximately a 300%–400% increase in sum rate over single-connectivity. This substantial improvement highlights the effectiveness of multi-connectivity in enhancing network capacity, particularly in environments where users are mobile, allowing users to benefit from aggregated gains. Such findings underscore the importance of connectivity type and beam angle optimization for maximizing throughput in VLC networks.

5 Conclusion

In this paper, the performance of single- and multi-connectivity systems has been examined in a novel communication-centric VLCS system equipped with MAP technology to tackle the link blockage and user mobility issues in VLC systems. To achieve this, non-convex optimization frameworks have been formulated to maximize the total sum rate for single- and multi-connectivity technologies by jointly optimizing MAP position and user-AP association while considering minimum data rate, sensing MI, and illumination requirements. The MM technique from the literature has been used to implement an algorithm that obtains a near-optimal solution for both single- and multi-connectivity frameworks. Simulation results have revealed that multi-connectivity systems significantly outperform single-connectivity frameworks in MAP-enabled VLCS networks. The results further revealed that both single- and multi-connectivity MAP-enabled VLCS networks outperform the fixed AP benchmark scheme, indicating improvements in total sum rate and mitigating the challenges in VLC systems. Future research directions include investigating the impact of interference in single- and multi-connectivity MAP-based VLCS networks and exploiting any mutual benefits of MAPs and ORISs in VLCS networks.

References

1. Aboagye, S., et al.: RIS-assisted visible light communication systems: a tutorial. *IEEE Commun. Surveys Tuts.* **25**(1), 251–288 (2023)
2. Pal, R.K., Dash, S.P., Aïssa, S., Ghose, D.: Outage probability analysis of RIS-assisted indoor wide-FOV-VLC/RF wireless communication. *IEEE Wireless Commun. Lett.* **13**(1), 34–38 (2024)
3. Ndjiongue, A.R., et al.: Re-configurable intelligent surface-based VLC receivers using tunable liquid-crystals: the concept. *J. Lightw. Technol.* **39**(10), 3193–3200 (2021)
4. Abdeljabar, S., Eltokhey, M.W., Alouini, M.-S.: Sum rate and fairness optimization in RIS-assisted VLC systems. *IEEE Open J. Commun. Soc.* **5**, 2555–2566 (2024)
5. Aboagye, S., et al.: Intelligent reflecting surface-aided indoor visible light communication systems. *IEEE Commun. Lett.* **25**(12), 3913–3917 (2021)
6. Hammadi, A.A., et al.: Deep Q-learning-based resource management in IRS-assisted VLC systems. *IEEE Trans. Mach. Learn. Commun. Netw.* **2**, 34–48 (2024)

7. Sun, S., Yang, F., Song, J.: Sum rate maximization for intelligent reflecting surface-aided visible light communications. *IEEE Commun. Lett.* **25**(11), 3619–3623 (2021)
8. Maraqa, O., Aboagye, S., Ngatched, T.: Optical STAR-RIS-aided VLC systems: RSMA versus NOMA. *IEEE Open J. Commun. Soc.* **5**, 430–441 (2024)
9. Saghir, Z., Aboagye, S.: Movable access point-aided integrated visible light communication and sensing networks. *IEEE Wireless Commun. Lett.* (2024)
10. Majamaa, M.: Toward multi-connectivity in beyond 5G non-terrestrial networks: challenges and possible solutions. *IEEE Commun. Mag.* 1–7 (2024)
11. Shi, D., et al.: Effect of illumination intensity on LED based visible light communication system. In: 2020 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), pp. 1–4 (2020)
12. Bettstetter, C., Hartenstein, H., Pérez-Costa, X.: Stochastic properties of the random waypoint mobility model. *Wireless Netw.* **10**(5), 555–567 (2004)
13. Soltani, M.D., et al.: Modeling the random orientation of mobile devices: measurement, analysis and LiFi use case. *IEEE Trans. Commun.* **67**(3), 2157–2172 (2019)
14. Aboagye, S., Ngatched, T., Dobre, O.A., Ibrahim, A.: Joint access point assignment and power allocation in multi-tier hybrid RF/VLC HetNets. *IEEE Trans. Wireless Commun.* **20**(10), 6329–6342 (2021)
15. Zhang, Q., et al.: Design and performance evaluation of joint sensing and communication integrated system for 5G mmWave enabled CAVs. *IEEE J. Sel. Topics Signal Process.* **15**(6), 1500–1514 (2021)
16. Sun, Y., Babu, P., Palomar, D.P.: Majorization-minimization algorithms in signal processing, communications, and machine learning. *IEEE Trans. Signal Process.* **65**(3), 794–816 (2017)



A Review on VLC-Based Indoor Positioning Technologies

Felipe Bueno^(✉) 

McMaster University, Hamilton, ON, Canada
buenof@mcmaster.ca

Abstract. Indoor positioning systems (IPS) are becoming increasingly important in the future of wireless communications, with applications ranging from smart agriculture to indoor location-based services. While numerous IPS solutions have been proposed in the literature, many of these technologies are either overly complex or lack the necessary accuracy. The emergence of visible light communication (VLC) technology has been pivotal in enabling more effective IPS. VLC-based systems can offer centimeter-level positioning accuracy at a relatively low cost, making them a promising solution for a variety of indoor applications. In this work, we provide a comprehensive review of the state-of-the-art in VLC-based indoor positioning systems. We examine the main positioning methods and provide examples in literature.

Keywords: Visible light communications · Indoor positioning systems · Wireless communications

1 Introduction

Applications such as smart agriculture [13], smart industries [27], smart hospitals, and location-based services demand highly precise positioning. While the Global Navigation Satellite System (GNSS) dominates as the main outdoor positioning setup with the global positioning system (GPS), it lacks positioning accuracy when it comes to indoor positioning tasks. Therefore, new technologies are necessary to provide a high accuracy Indoor Positioning System (IPS).

Infrared (IR), ultrasound, radio-frequency identification (RFID), wireless local area network (WLAN), Bluetooth, sensor networks, and ultra-wideband (UWB) are among many possible radio frequency technologies that have been used to realize IPS. For example, in [8], the authors propose an IR-based IPS, achieving accuracy results on the order of centimeters with a mean positioning error of 6 cm. In [3], an experiment where IPS schemes based on Wi-Fi and Bluetooth low energy (BLE) is described. The experiments carried out demonstrate an average positioning error of 1.3m and 1.7m for Wi-Fi and BLE, respectively. RFID-based systems are also proposed in [24, 25]. In [24], an RFID reader interrogation pulse is employed to synchronize tags included in the emitter and sensor circuitries. The sensor's position is estimated geometrically based on distance measurements. Experimental results indicate an accuracy value below 5 cm

with a positioning error of 15 cm in the worst case. In [25], an off-line trained machine-learning model is employed to estimate the time of arrival (TOA) of an ultrasonic signal. The required synchronization between the emitter and sensor is performed using the RFID standard protocol features. The experiments demonstrate an accuracy of 25 cm in the worst case. Although these approaches are capable of achieving high accuracy with small errors, most of these approaches are prone to electromagnetic interference, which are prohibited in certain environments or need extra infrastructure to be implemented [23].

Visible light communication (VLC) systems are emerging as one of the most promising technologies for the sixth generation (6G) of mobile communication. Compared to alternative options, VLC systems offer advantages such as low cost, security, and a much larger available bandwidth than radio-frequency systems. VLC is also a key enabler for indoor positioning systems (IPS), as it can provide centimeter-level accurate positioning. VLC-based IPS, also known as visible light positioning (VLP), has gained attention due to its low cost and high accuracy [6]. Moreover, commercial off-the-shelf LED lamps that consume low power, have a lifespan of up to ten years, and entail low maintenance costs, can be employed as typical VLP transmitters without generating radio interference [6].

There are three main Visible Light Positioning (VLP) methods as illustrated in Fig. 1: the mathematical approach, which is subdivided into three categories - proximity, triangulation, and fingerprinting; the sensor-assisted approach, which employs sensors such as image detectors, accelerometers, and multiple optical receivers to aid in the receiver's positioning; and the position optimization approach, which uses filtering techniques such as Kalman filter (KF) and particle filter (PF) algorithms, spring model-based methods, and normalization methods.

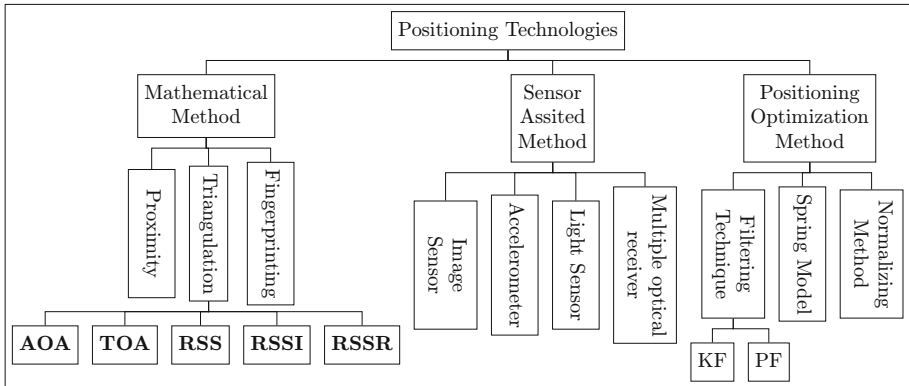


Fig. 1. Taxonomy of positioning technologies [23].

In this work, we present an overview of the current state-of-art of VLP technology. First we discuss a few requirements of a VLP-based system. Next we provide details on the main positioning methods, the mathematical, sensor-assisted

and through optimization techniques. Finally, we review a few of the most recent papers published in VLP systems.

2 VLC-Based Positioning

The primary benefit of a VLP system lies in its ability to seamlessly integrate with existing lighting infrastructure [6]. Consequently, such a system must adhere to a range of illumination, compatibility, and considerations for human health impacts as delineated by regulatory standards.

The illumination requirements cover aspects such as maintained illuminance, uniformity of illumination, glare rating, and color rendering index. The maintained illuminance establishes a lower bound for the average illuminance over the course of the lighting infrastructure's lifetime [6]. The criterion of uniformity of illumination sets a lower limit for the ratio of the minimum and mean illuminance on a surface, in order to avoid eye discomfort due to constant vision adaptation [6]. The glare criteria establish limits to excessive illumination that can potentially cause temporary blindness, and color rendition regulates the minimum color rendering index, which is an indicator of the extent to which the lamp enables a natural color gamut perception by an observer [6].

Regarding compatibility, the VLP system should be able to integrate with the existing infrastructure. This criteria includes the VLP system being able to smoothly connect with pre-existing control systems such as the Digital Addressable Lighting Interface (DALI) and DMX512 [6]. Furthermore, it should also be able to integrate with dimming systems, implemented either with Pulse-Width Modulation (PWM) or Amplitude Modulation (AM) [6].

Moreover, VLP systems should consider the impact of modulation on health or perception. It is known that temporal light artifacts, such as flickering, can worsen medical disorders such as photosensitive epilepsy, migraine, nausea, and asthenopia (eye strain) [7, 18, 36]. Hence, a VLP system must implement safe frequency and depth modulation. For instance, the IEEE 1789:2015 standard, [1], introduces guidelines for modulating current in high-brightness environments to mitigate potential adverse health effects.

When determining these criteria for different environments, considerations extend beyond mere compliance to standards. Psycho-physiological factors such as visual comfort, safety concerns, ergonomic considerations, task requirements, and economic implications are all taken into account [6]. Notably, standards also encompass more intricate parameters, metrics (e.g., cylindrical illuminance), and guiding [6].

The block diagram represented in Fig. 2 illustrates the building blocks of a VLP/VLC system [6]. The blocks with a white background represent elements that are not specific to VLC/VLP and can be borrowed from RF systems. In contrast, the modulating, transmitting, and receiving elements, depicted in the blocks with a blue background, require dedicated components. One of the significant disparities compared to RF is that the transceiver combination only supports modulating with positive and real-valued signals [6].

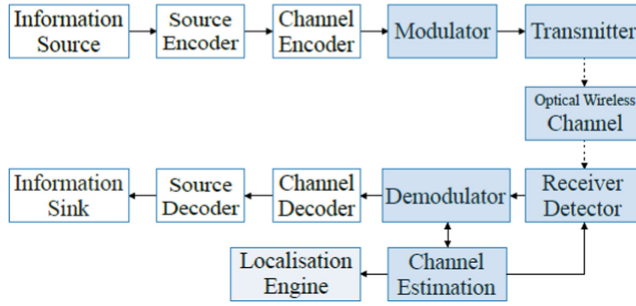


Fig. 2. Block diagram representation of the VLP/VLC building blocks [6].

In the following sections we describe the methods for finding the position of a mobile device through a VLC system. We begin by introducing the mathematical method, next we explore the sensor-assisted method and finally we extend our discussion to the position optimization method.

3 Mathematical Method

The mathematical approach encompasses the proximity, fingerprinting, and triangulation techniques. The most basic of these methods is the proximity approach, where a device's position is inferred by knowing the position of the beacon with the strongest signal. The fingerprinting method leverages from features of signals that are location dependent, such as received signal strength (RSS). Finally, in the triangulation method, the position of the device can be estimated through the geometric properties of triangles.

3.1 Proximity Method

When two or more beacons can detect a mobile device, the device's position can be estimated by mapping it to the position of the beacon with strongest signal. However, the accuracy of the proximity method relies on the quantity of beacons present, making it non-scalable. Additionally, the proximity method's variance is too significant to meet positioning requirements effectively. Nonetheless, the proximity method has been widely explored in the literature, presenting relatively good results.

In [37], the authors introduce a LED-Identification (LED-ID) detection and recognition method utilizing visible light positioning via proximity techniques. The method involves manipulating the frequency and duty cycle of a Pulse-width Modulation (PWM) signal to impart unique characteristics to the LED-ID. Subsequently, Complementary metal-oxide-semiconductor (CMOS) image sensors capture images of the LEDs, and an offline-trained a Fisher classifier and a linear support vector machine (SVM) are used to identify the LED-ID with the strongest signal. Once the LED is correctly identified, the position of the

CMOS receiver is estimated by proximity. In the experimental tests carried out, 1035 LEDs could only be labeled using a PWM signal. All LEDs were correctly identified using the SVM and Fisher classifiers, resulting in a positioning error of a maximum of 6 m. Although, the accuracy is not at the level of centimeters, it is sufficient for large-scale applications.

In [35], a VLP system that integrates proximity with Pedestrian Dead Reckoning (PDR) is introduced. PDR relies on the inertial measurement unit (IMU) to estimate location. Because PDR determines position based on step count, its accuracy depends on accurately measuring pedestrian step size and heading direction, which can be challenging. As a result, a significant drawback of PDR technology is the accumulation of measurement errors. To calibrate the PDR and prevent error accumulation, the authors of [35] propose using VLP. PDR addresses the discontinuity issue caused by the limited field of view (FOV) of camera sensors, while VLP provides global data and periodic calibration for PDR. The integrated system location algorithm is illustrated in the schematic shown in Fig. 3. Experimental results indicate an accuracy of 98.5%.

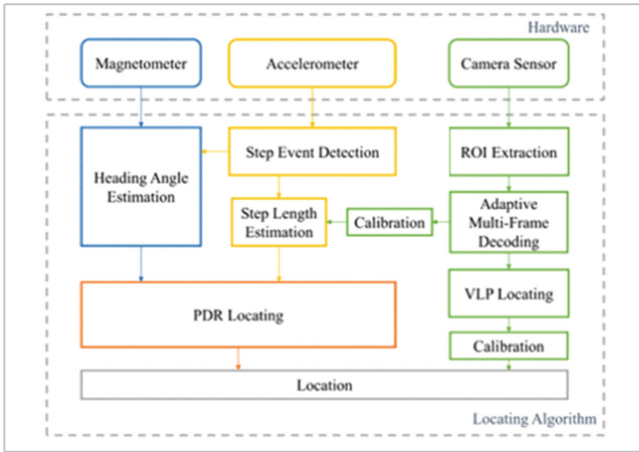


Fig. 3. VLP-integrated PDR system [35].

3.2 Fingerprinting Method

In the fingerprinting method, a device's position is estimated by measuring signal features or characteristics that vary with location. Fingerprints are categorized into two types: radio-map-based and map-free. Map-based fingerprinting localization consists of two stages: offline and online (or run-time), as illustrated in Fig. 4. The offline stage involves conducting a site survey indoors to create a radio map. The location positioning technique estimates a target's position based

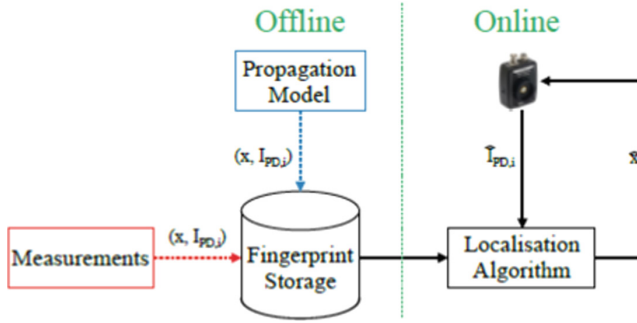


Fig. 4. Schematic of the manual (red) and model-based (blue) fingerprinting process [6] (Color figure online).

on observed signal intensities and the radio map. Diffraction and reflection can affect RSS, leading to low accuracy [23].

In [10], a VLP system based on the weighted K-nearest neighbor (WKNN) fingerprinting method is proposed. In this approach, position coordinates are mapped to an indicative parameter, the Received Signal Strength Indication (RSSI). As a preprocessing step to mitigate the effects of random fluctuations, multiple RSSI measurements for each sampling point are collected and averaged. The WKNN algorithm is then trained offline using the processed data. Experimental results demonstrate a minimum positioning error on the order of centimeters.

In [38], to combat the effects of the errors introduced by non-uniform and low-density location fingerprinting distribution, a VLP method based on location fingerprinting and meta-heuristic is proposed. In this method, all possible cases of uneven distribution of location fingerprinting are considered by introducing the random selection rate. The illustration of the proposed fingerprinting positioning model where M LEDs are placed at the ceiling is shown in Fig. 5. The grid is divided in a total of N fingerprinting points minimum spacing of S . During the off-line phase, the target is placed at the N fingerprinting points and the RSS values from the M LEDs are collected. To improve location accuracy, the fingerprinting point weights are optimized through a series of meta-heuristic algorithms and the results are compared. The simulation results indicated an average positioning error of 3.11 cm.

Map-based fingerprinting is a complex approach, as measurements from multiple site positions must be meticulously collected [6]. Furthermore, positioning accuracy can be compromised by environmental changes that differ from the data used for training [6]. Map-free fingerprinting methods can simplify the process of maintaining location mappings.

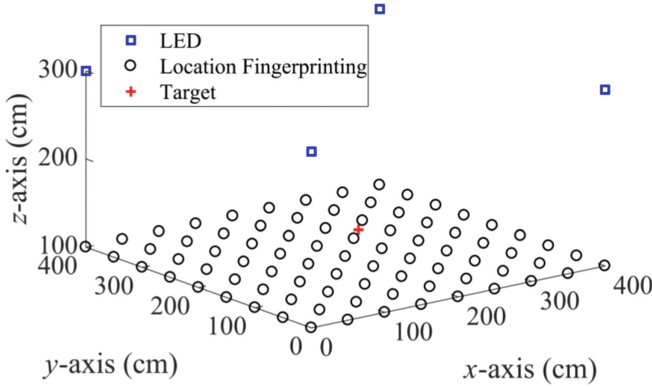


Fig. 5. Schematic of the manual (red) and model-based (blue) fingerprinting process [38] (Color figure online).

3.3 Triangulation

The position of a mobile device can also be estimated through geometric properties of triangles, a process called triangulation. This method requires the measurement of distances to a fixed point or the angle-of-arrival (AOA) of a signal. The three main types of triangulation methods are AOA, Time-of-Arrival (TOA), and received signal strength (RSS). Each triangulation method is explained below.

a) AOA: The AOA method determines the device's position by analyzing the angle between the base station and the mobile device. Employing the AOA technique requires a minimum of two base stations. By obtaining two or more angles of incidence, it is feasible to calculate the intersection of multiple lines using geometric principles. This intersection denotes the mobile device's location, as shown in Fig. 6. This positioning concept relies on utilizing the convergence of multiple directions of the wireless signal to estimate the mobile device's location [23].

Given the mobile device's coordinates as (x_0, y_0) and the base station's as (x_i, y_i) , and if θ_1 and θ_2 denote the angles formed by the wireless signal's azimuth line and the base stations (BS1 and BS2), then they satisfy the following mathematical expression [23]:

$$\tan \theta_i = \frac{y_i - y_0}{x_i - x_0}, \text{ with } i \in \{1, 2\} \quad (1)$$

The precision of AOA in determining location is inferior to that of TOA, and this approach is highly influenced by the external environment. Furthermore, as the distance between the mobile device and the base station increases, AOA's accuracy diminishes [23]. Additionally, implementing AOA requires extra hardware, thereby raising the overall cost [23].

A few triangulation examples using AOA are presented next. In [15], the authors introduce a VLP system using quadrant-solar-cell (QSC) and third-

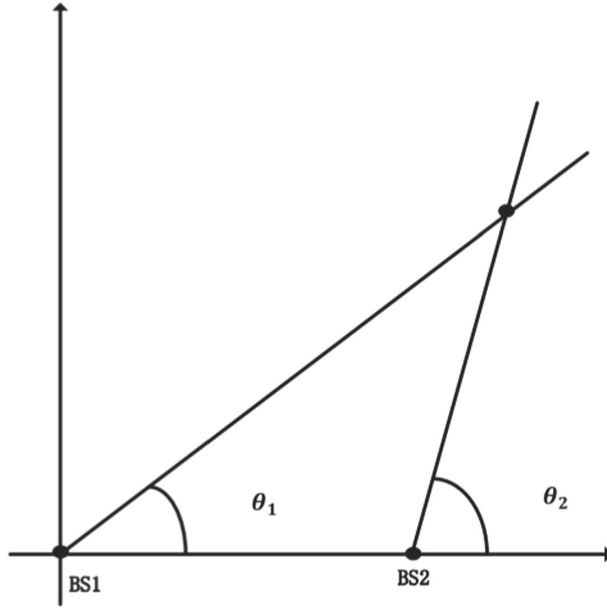


Fig. 6. Positioning based on AOA [23].

order ridge regression machine learning (RRML). The investigated QSC is used as an angular diversity receiver. The proposed QSC is an integrated device made up of 4 identical crystalline silicon (Si) solar cells. The quadrants of the QSC are separated by barriers shown in Fig. 7 (a). These barriers can change the amount of received light when the light source is located at different angles with respect to the QSC. Position information can be derived by analyzing the photo-currents received at the four quadrants of the QSC. The experimental results demonstrate a positioning error of 3.0881 cm.

In [28], angular diversity is achieved through the design of a photodiode (PD) receiver's aperture. The authors propose an algorithm based on the maximum likelihood (ML) principle to estimate the AOA. By triangulating the estimated AOA, they determine the receiver's position in 3D. The results demonstrate an accuracy of 10 cm or better, even with a limited number of LEDs, such as four LEDs for a $5\text{ m} \times 5\text{ m}$ area.

A visible light positioning (VLP) system using a two-step approach based on the angle of arrival (AOA) is considered in [29]. In the first step, the AOA is estimated based on the received optical signals. In the second step, the receiver's position is determined using triangulation. The angles are estimated by measuring the differences in the RSS values at different PDs of the receiver. To evaluate the accuracy of angle estimation, the authors use the Cramer-Rao lower bound (CRLB). The results indicate a positioning accuracy in the order of centimeters.

b) TOA: While in the AOA approach the device's position is estimated based on the angle between the base station and the mobile device, in the TOA app-

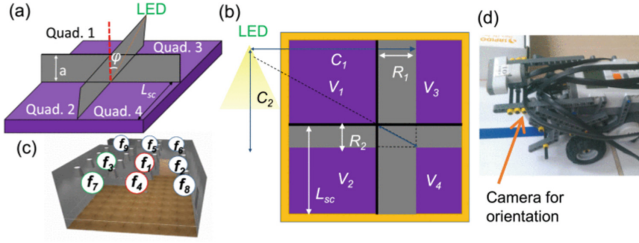


Fig. 7. (a) Principle of QSC for angular diversity AOA VLP Rx. (b) Top view of the QSC with different shadows produced by the barriers. (c) Unit cells with different beacon frequencies for whole area positioning. (d) One example of installing QSC on a robotic trolley [15].

reach the device's position is estimated based on the time it takes for the signal to reach the base station [23].

Assuming that v is the propagation speed and t_i is the propagation time between the mobile device and the i -th base station. Then, $v \times t_i$ can be used to estimate the distance s_i between the mobile device and base station. Assume that the base station's coordinates are (x_i, y_i) and the mobile device's coordinates are (x_0, y_0) . After that, we may calculate the following formula:

$$(x_i - x_0)^2 + (y_i - y_0)^2 = s_i^2, \text{ with } i \in \{1, 2, \dots, n\} \quad (2)$$

where n is the number of detected base stations. In the ideal case, the spatial position of the device can be localized in the intersection of three circles whose centers are the base station and the radius are $s_1, s_2, \dots, s_n$ as shown in Fig. 8 [23].

One major drawback of the TOA method is the requirement for accurate time synchronization between the transmitter and receiver [23]. Therefore, the positioning is highly dependent on hardware, which might increase the costs of implementation [23].

A VLP system based on the TOA approach has been presented in [33]. The transmitter and receiver are assumed to be perfectly timely synchronized. As illustrated in Fig. 9, LEDs are employed in this system for both signal transmission and lighting. After receiving the signals, the device (receiver) calculates each LED's distance. The LED's TOA signal can be measured to determine the range. The Cramer-Rao bound is used to examine the accuracy of the TOA estimation.

A Hybrid TOA/RSS-based distance estimate for VLP systems is proposed in [19], and its Cramer-Rao lower limit is analytically compared against the Cramer-Rao lower bounds of TOA-based and RSS-based distance estimation. Asymptotic results are discussed in the paper.

A possible solution for the strict time synchronization requirement introduced by the TOA method is the Time Difference of Arrival (TDOA) approach. In this approach, the propagation time difference between the mobile device and base

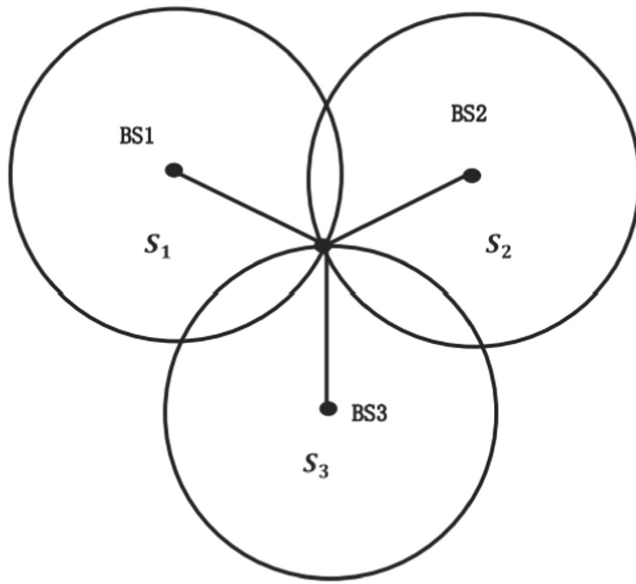


Fig. 8. Positioning based on TOA [23].

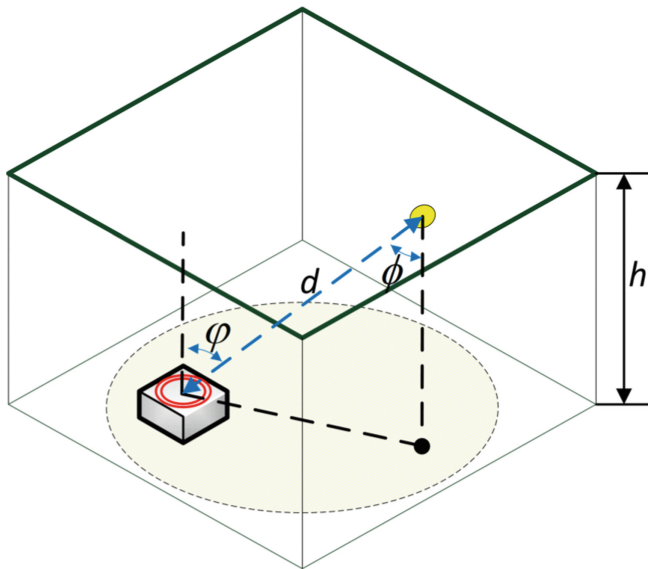


Fig. 9. Schematic diagram of the positioning system for indoor localization presented in [33].

stations is measured and then converted into the distance by multiplying the speed of propagation. The main advantage of the TDOA method is that accurate time synchronization between the base stations and the target is not necessary; only time synchronization between the base stations involved in positioning is required.

The location of the mobile device can be attained by the hyperbolic positioning method. The distance difference between base stations BS1 and BS2 is $s_{12} = v \times (t_1 - t_2)$, as Fig. 10 illustrates. The mobile device must fall inside the hyperbola, with the two base stations serving as its focuses and the distance difference (s_{12}) as the focal length (see to Fig. 10's solid line). In the meantime, other hyperboles, represented by the dotted lines in Fig. 10, can be created, using the BS1 and BS3 as its foci and their distance difference, s_{13} , as the focal length [23]. The mobile device is then located at the intersection of the two hyperboles. Assume that the base station's coordinates are (x_1, y_1) and the mobile device's coordinates are (x_0, y_0) , then we can get the following equation [23].

$$s_{li} = \left(\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} - \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} \right), \text{ with } i \in \{1, 2\}. \quad (3)$$

Equation (3), yields two results representing the two hyperbola intersections. When applying it to real applications, it necessary prior knowledge to distinguish which is the real-world coordinates.

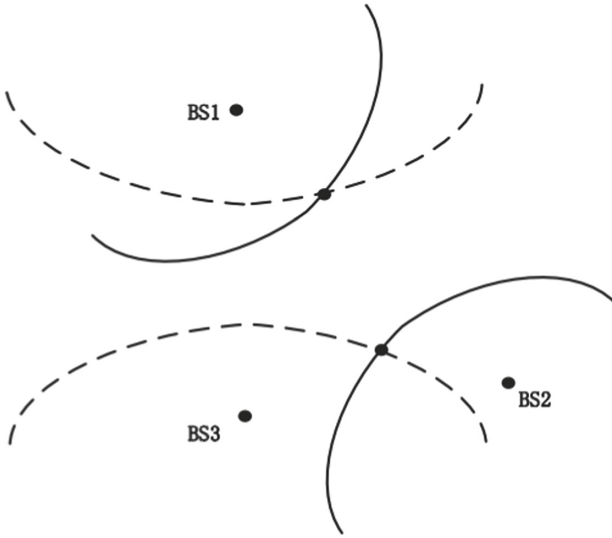


Fig. 10. Positioning based on TDOA [23].

A few example of publications involving TDOA are presented next. In [11], an indoor VLP system using the TDOA method is proposed. The proposed

system requires no ID information about the LED panel. Additionally, the proposed scheme utilizes cheap photodiodes as receivers. The simulation results indicates an positioning accuracy of 3.9 cm with clock precision of 1 ns. In [12], a low-complexity TDOA-based indoor VLP system using an enhanced practical localization scheme depicted in Fig. 11 is proposed and experimentally demonstrated. This method adopts the TDOA measurement scheme by means of cross-correlation armed with the cubic spline interpolation.

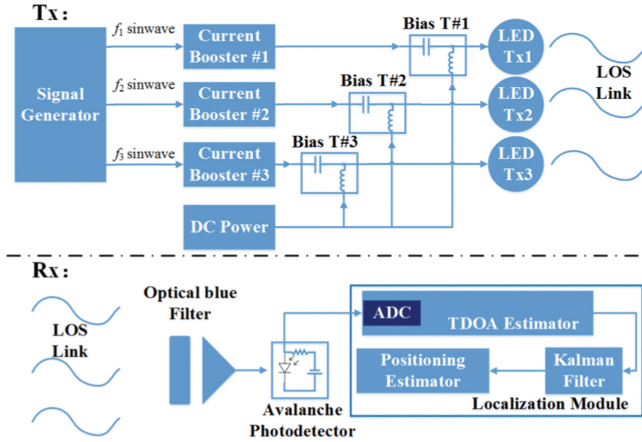


Fig. 11. Positioning based on TDOA [12].

Moreover, the transmitter synchronizes carrier waves across all channels with a signal generator and amplifies them, powering each LED with its respective amplified carrier. The receiver comprises of an avalanche photodetector (APD) and a localization module. The APD captures mixed RF signals from LED transmitters, processed by the localization module. To enhance accuracy, an optical blue filter on the APD eliminates slow fluorescent elements. The localization module functions in three stages: TDOA measurement via cross-correlation of RF carriers, error reduction employing Kalman filtering, and position estimation using trilateration. The experimental results show an average positioning accuracy of 9.2 cm.

c) RSS: In RSS-based indoor VLP systems, the estimation of the position of a receiver is based on RSS measurements between the receiver and transmitters [23]. This is possible, due to the correlation of the distance from the transmitter to the receiver, and the attenuation in the received signal strength.

Typically, multipath propagation is a major obstacle to achieving accuracy in RSS-based indoor positioning systems. This occurs because only the part of the signal that corresponds to the line of sight (LOS) is useful for calculating the distance between the transmitter and receiver [23]. Therefore, in [30], a RSS-based indoor VLP method using channel state information (CSI) is proposed.

To minimize the effects of the multipath propagation and improve accuracy, the author employ CSI to estimate the normalized amplitude of the LOS path. With the knowledge of the LOS path received power, the distance between transmitter and receiver can be better estimated. The simulations showed that the proposed method can pinpoint a location within 6.1 cm, with a maximum error of 17.7 cm. Compared to the traditional least-squares (LS) method, the proposed system offer 83% reduction in average error and an 81% decrease in root mean square error (RMSE).

In [9], a RSS-based indoor VLP system is introduced. The authors propose a system model as depicted in Fig. 12. A joint maximum-likelihood approach is used estimate to receiver horizontal coordinate and receiver characteristic vector based on the RSS information. The proposed solution is robust in respect with the receiving coefficient and receiving angle. In fact, no receiving angle information is required. Besides, the experimental results demonstrate that the proposed method achieves positioning accuracy below 7 cm.

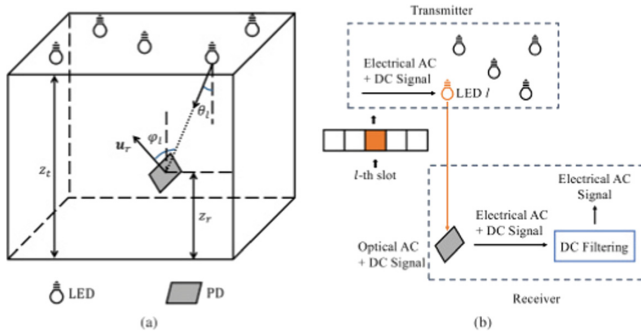


Fig. 12. Illustration of the indoor VLP system presented in [9]. (a) System configuration [9]. (b) Signal flow [9].

4 Sensor Assisted Method

To improve the localization accuracy, multiple types of sensor can be used to assist a particular positioning method. These includes the use of image sensors, accelerometers , light sensors, and multiple optical receivers.

1) *Image Sensor*: The main idea behind the use of image sensors as an assisting device in VLP system is to capture images and process it in order to identify luminaries, decode their identifiers, when existing, and estimate position or distances.

In general, receiver devices are subject to rotation about the x-, y- or z-axis, which may be caused by human handling, for instance. This can impact the performance of the image processing algorithms necessary for extracting useful information from image sensors. Therefore, in [31], a tilted receiver camera

correction and partially blocked LED image compensation algorithm to realize high-accuracy positioning for VLP systems is proposed. Figure 13 illustrates the schematic of the proposed indoor visible light positioning (VLP) system utilizing a single circular LED. LEDs transmit their unique digital IDs, containing embedded positional data stored in a central database, through continuous modulation. A user captures images of a LED periodically using a smartphone's front camera. The system then performs two key tasks. Firstly, it decodes the rolling shutter patterns present in the captured images to extract the LED's unique ID and subsequently retrieve its corresponding world coordinates from the database. Secondly, the system detects the LED's boundary within the images and calculates the camera's translation matrix. Smartphone sensor data is leveraged to estimate tilt angles, enabling the computation of the camera's rotation matrix. Finally, the system determines the camera's world coordinates. The experimental findings demonstrate that the suggested VLP approach can attain an average positioning error of less than 7 cm when capturing over 40% of the LED image, and a positioning error averaging 3.9 cm when capturing over 90% of the LED image.

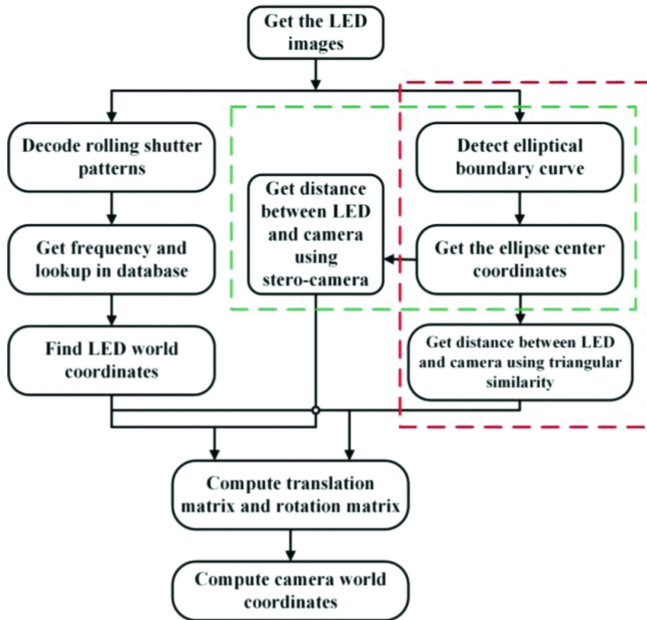


Fig. 13. Tilted receiver camera correction method for a single-LED-based VLP system proposed in [31]: (1) finding the LED's world coordinates, (2) computing the camera matrices.

In [16], a non-line-of-sight (NLOS) VLP system assisted with a image sensor is proposed. The authors consider an indoor environment where a LED is

illuminating a room and a COMS camera is used for capturing the reflected modulated lights from the floor. In a first stage, a short camera burst captures the light signal from the LEDs. This information is used to calculate the location of the LEDs. Then, a longer camera exposure captures a detailed image of the reflected light. By analyzing the shapes of light spots and a reference label on the ground, the system estimates the camera's position and orientation. The experimental results indicates that the proposed method is capable of achieving 90-th percentile accuracy of less than 22 cm for three-dimensional (3D) positioning.

2) *Accelerometers*: An accelerometer, Fig. 14, is a device that senses tilt and therefore can be used to determine the orientation of a receiver, and consequently, the angle of incidence of the signal. This information is valuable in techniques like RSS-based positioning, where how the device is held affects how well it finds its location.

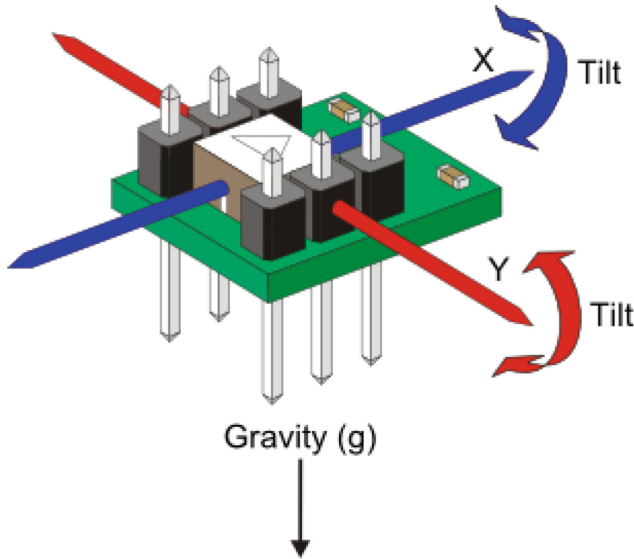


Fig. 14. Accelerometer sensor [17].

In [17], an accelerometer-aided visible light positioning (VLP) system that enables mobile device localization, even with a tilted camera, is introduced. The system is comprised of four ceiling-mounted LEDs, each transmitting its unique 3D world coordinates. An image sensor, at an unknown position, receives and demodulates these signals. The 3D position of the mobile device is determined based on the LEDs' broadcasted 3D coordinates and the captured 2D image coordinates. The proposed algorithm achieves this localization in two steps. First, it determines the distances between the image sensor and each LED using a single image sensor, unlike conventional approaches requiring two. Second, these distances are incorporated into a set of equations, which are then solved to yield

the image sensor's position. Notably, the inclusion of accelerometer data allows the image sensor to operate at any angle, enhancing system robustness.

In [41], a positioning system utilizes an accelerometer to measure a receiver's acceleration, which then helps to determine the receiver's orientation. By combining the received signal strength and accelerometer data, the system can estimate the receiver's position through a mathematical formula. Additionally, the accelerometer measurements eliminate the need to assume that the transmitter's irradiance angle matches the receiver's incidence angle. This eliminates the requirement for the receiver's height from the ground. Experiments in a simulated environment demonstrated an accuracy of about 0.5 m.

3) *Light Sensor*: A light sensor has sensitivity to both distance and incident angle of the light signal. This characteristic makes them a suitable device to be used as a receiver. The light sensor can retrieve the beacon information, measure RSS and then estimate the position of the mobile device.

In [14], a passive VLP systems using light sensors is proposed. This system, depicted in Fig. 15, achieves target localization by analyzing variations in the RSS of ambient light. This approach eliminates the requirement for active devices or tags on objects of interest and avoids the need for modifications to existing lighting infrastructure. Experimental results within a $2\text{ m} \times 3.6\text{ m}$ testbed demonstrate a median localization error of 7 cm and a 90-th percentile error of 21 cm.

A method for the locating a smartphone by receiving sinusoidally modulated light signals emitted by ceiling-mounted LEDs is proposed in [2]. The proposed VLP method works by measuring the reflected illumination from the floor through an ambient light sensor mounted on the smartphone. The experiments using four LEDs modulated at a frequency of about 60 Hz and a highly reflective floor material showed that the 90th-percentile positioning error was 35.5 cm.

4) *Multiple optical receivers*: The mathematical methods usually require multiple transmitters or receivers to estimate angles and/or distances. However, an important issue that must be taken into account is Inter-Cell Interference (ICI), which in VLP systems occurs when signals from multiple LEDs in the same room can interfere with each other. To avoid ICI techniques that employ a single transmitter and multiple receivers can be used.

In [39], a VLP system, illustrated in Fig. 16, using a single transmitter, multiple receivers and RSSI technology is proposed. If one receiver is selected as a reference, say Rx_1 , the relative positions of Rx_2 and Rx_3 can be expressed as

$$\begin{aligned} Rx_1 &= (r \cos(\varphi), r \sin(\varphi)) \\ Rx_2 &= (r \cos(\varphi + 2\pi/3), r \sin(\varphi + 2\pi/3)) \\ Rx_3 &= (r \cos(\varphi + 4\pi/3), r \sin(\varphi + 4\pi/3)) \end{aligned} \quad (4)$$

where φ is the device's rotation angle, Rx_i is the i -th receiver, and r is the radius of the device. Knowing the distance between the LED and the receiver, along with the LED's location and the receiver's specifications, allows us to estimate

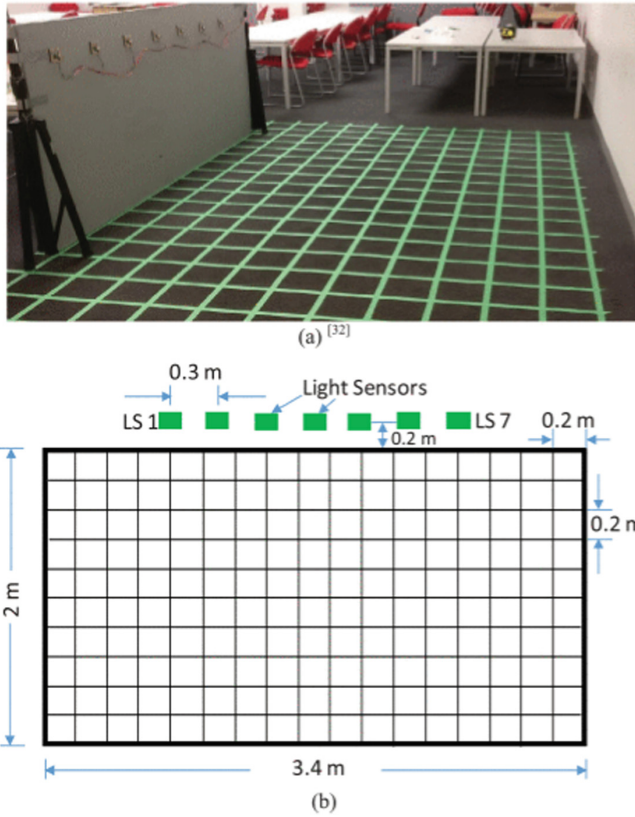


Fig. 15. Proof of concept system. (a) Actual setup. (b) Layout diagram. [14].

the position of the receiver. The results indicate that a positioning error of 1.5 cm is achievable.

The authors of [40] propose a 3-D positioning algorithm based on gain difference, which is a function of the angle of arrival and the received signal strength. The system is comprised of a single transmitter and multiple tilted receivers to avoid ICI. The experimental results indicate a positioning error of less than 6 cm.

5 Positioning Optimization Method

Although mathematical and sensor-based methods achieves a high accuracy. There is still room for improvements through advanced techniques that includes the use of Kalman and particle filters, spring model and normalization methods.

1) *Kalman filter*: The Kalman filter is a powerful tool used to estimate the state of a system, even when the measurements you have are imprecise or uncertain. This is particularly important for situations where things are constantly changing such as in indoor positioning tasks. Although the Kalman filter requires

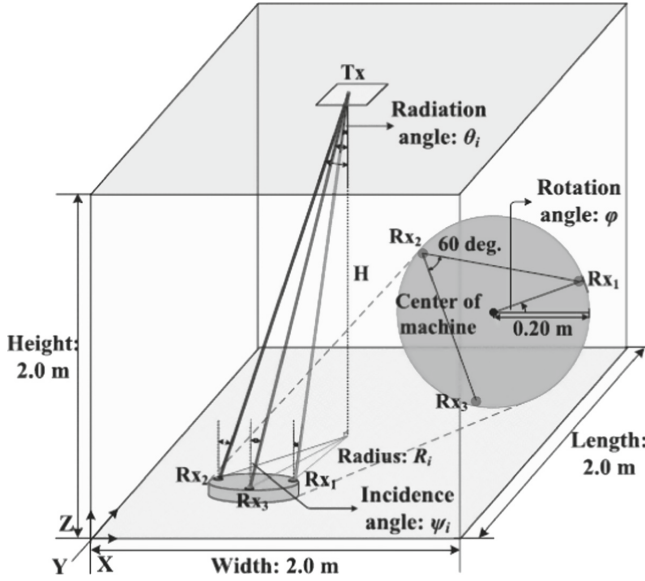


Fig. 16. Positioning system based on single transmitter and multiple receivers proposed in [39].

a dynamic and linear model, its enhanced version, the extended Kalman Filter (EKF) is able to linearize a nonlinear model.

As an alternative to EKF, there is the unscented Kalman filter. While EKF treats the non-linearity using analytical linearization, the unscented Kalman Filter performs statistical linearization based on a set of rules [5]. In [22], the unscented Kalman filter is used in combination with VLP technology and inertial navigation data to decrease the error caused by the multipath reflection effect. The experimental results demonstrate an accuracy in the centimeter order.

2) *Particle filter*: Particle filters, also known as sequential Monte Carlo methods, are a powerful tool for estimating the state of a system in real-time. Unlike the Kalman filter, which is limited to linear systems with Gaussian noise, particle filters can handle non-linear and non-Gaussian scenarios. Particle filters employ a set of weighted samples, also referred to as particles, to represent the state of the system. These particles are effective in handling non-linear state-space models and arbitrary initial state and noise distributions. Notably, particle filtering techniques offer a robust methodology for generating samples from the target distribution without imposing restrictive assumptions on the underlying dynamics or the state distributions. However, the computational complexity of particle filters can become significant for high-dimensional systems.

In [32], a novel visible light localization (LiLoc) algorithm, based on augmented particle filter algorithm leverages from ambient light, without requiring light source modulation. LiLoc directly utilizes the fingerprinting of ambient light intensity (e.g., fluorescent, incandescent, LED) to estimate target location.

To enhance accuracy, a light-intensity trajectory matching scheme refines target position. Additionally, a beacon identification algorithm, capitalizing on inherent visual features of light sources, is introduced. This algorithm, combined with PDR, provides positioning in challenging conditions (sunlight interference, sparse beacons, beacon damage, or missing training data). To optimize power consumption, an adaptive scheduling mechanism activates appropriate algorithms based on sunlight detection. The experimental results demonstrate that LiLoc achieves a median localization error of 2.2 m in complex lighting environments with diverse light sources and sunlight interference.

3) *Spring model*: Localization Algorithm based on a Spring Model (LASM) leverages the principles of a physical spring system to estimate node positions. Sensor nodes are modeled as particles interconnected by virtual springs to neighboring nodes. As a familiar concept, springs naturally exert forces that return displaced objects to their equilibrium positions. By calculating these inter-nodal forces within the LASM framework, the locations of unknown nodes (often termed "blind nodes") can be inferred [23].

4) *Normalizing method*: Estimations might introduce large errors, therefore normalization can be introduced to approximate estimated values, such as estimated distances, to the real values. For example, in [20] a normalizing method is used to approximate estimated distances to real ones. The degree of the distance adjustment can be manipulated by adjusting a controlling parameter.

6 Current State of the Art

In this section we present some examples of the most recent works published on VLP systems.

The system proposed in [26] combines visible light, infrared (IR) light and powerline communications to provide LED-based VLP functions within a building and support infrastructure-wide data communications for large-scale smart hospital operations. The experimental results indicated an accuracy within 10 cm.

In [4], a VLP system specifically designed for seamless indoor navigation within museums is proposed. The system leverages a smartphone's IMU to combine PDR with an AOA-based positioning algorithm. In standalone operation, the VLP AOA algorithm demonstrates high accuracy with a mean error of 10.64 cm. However, the PDR algorithm, operating independently, suffers from environmental influences resulting in a mean error of 3 m. Integration of both algorithms proves critical to mitigate these cumulative errors. By dynamically correcting PDR drift with the high-precision VLP AoA solution, the overall system significantly reduces its mean error to 0.85 m.

In [34], the authors employ power over Ethernet (PoE), which can transmit both energy and information simultaneously over the same cable, to support LED-based illumination and wireless information services. However, PoE-enabled VLP networks have issues such as low modulation bandwidth, low sampling rate under the resource-limitation case, and strict synchronization requirements. Specifically, synchronization between transmitters and synchronization

between transceivers are required to ensure the performance of the PoE-enabled VLP network. Therefore, in the proposed scheme, Fig. 18, the PoE structure is used to transmit the energy and both positioning and dimming control information for each LED. Moreover, the beacon is designed to unlock the time synchronization between the transceivers. The experimental results demonstrate a positioning accuracy in the order of the centimeters (Fig. 17).

In [21], the authors investigate a visible light communication (VLC) system incorporating intelligent reflecting surfaces (IRS). Their proposed model involves a VLC receiver gathering signals from multiple LED transmitters. These signals can travel directly via line-of-sight (LOS) paths, or they can bounce off IRSs before reaching the receiver. The authors also propose a method to optimize the adjustments of the IRSs to maximize the received signal strength. This method relies on mathematical formulas and considers situations where LOS paths are blocked. To evaluate its effectiveness, a benchmark is established using a generic

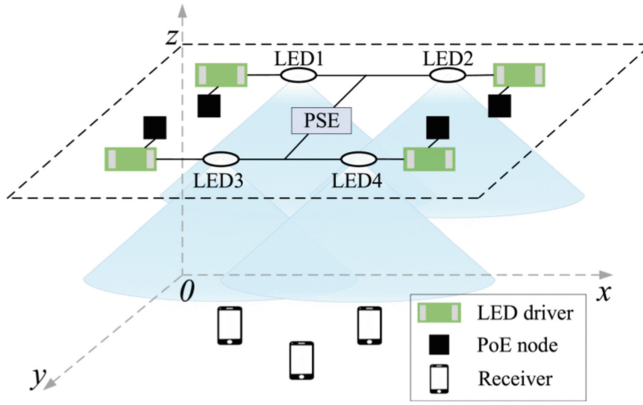


Fig. 17. PoE-enabled VLP network structure presented in [34].

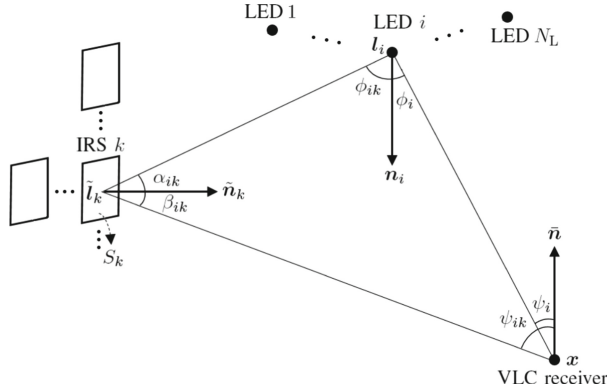


Fig. 18. PoE-enabled VLP network structure presented in [21].

Cramer-Rao Lower bound expression. Finally, the results indicate that while IRSs do not provide critical improvements in positioning accuracy with LOS signals from enough LED transmitters, they become very important for accurate positioning when most or all LOS paths are blocked.

7 Final Remarks

The visible light positioning technology is of paramount importance for the next generation of mobile communications technology. Since it is free from electromagnetic interference, it can be used in places where the use of RF could be prohibited, such as hospitals. Additionally, the larger bandwidth of VLC-based systems contributes to a higher data rate. In this work, an overview of VLP systems and their current state of the art was presented. Besides discussing the implications of adopting a technology primarily designed for illumination in wireless communications and positioning, the main positioning methods in VLC were also explored. These methods include mathematical approaches consisting of the *proximity*, *fingerprinting*, and the *triangulation* methods, of types AOA, TOA, TDOA, and RSS. Sensor-assisted methods using image sensors, accelerometers, light sensors, and multiple optical receivers, as well as optimization-based techniques such as Kalman and particle filters, were also discussed. Finally, a few of the most recent publications on indoor visible light positioning were presented.

References

1. 1789-2015 IEEE Recommended practices for modulating current in high-brightness LEDs for mitigating health risks to viewers, pp. 1–80. IEEE (2015)
2. Abe, K., Sato, T., Watanabe, H., Hashizume, H., Sugimoto, M.: Smartphone positioning using an ambient light sensor and reflected visible light. In: 2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–8 (2021)
3. Abed, F.A., Hamza, Z.A., Mosleh, M.F.: Indoor positioning system based on wi-fi and bluetooth low energy. In: 2022 8th International Engineering Conference on Sustainable Technology and Development (IEC), pp. 136–141 (2022)
4. Alcázar-Fernández, A., De-La-Llana-Calvo, Á., Lázaro-Galilea, J.L., Pérez-Navarro, A., Gil-Vera, R., Gardel-Vicente, A.: Seamless mobile indoor navigation with VLP-PDR. *IEEE Sens. J.* **24**(7), 11504–11514 (2024)
5. Unscented Kalman filter. <https://www.kalmanfilter.net/ukf.html>. Accessed 23 April 2024
6. Bastiaens, S., Alijani, M., Joseph, W., Plets, D.: Visible light positioning as a next-generation indoor positioning technology: a tutorial. *IEEE Commun. Surv. Tutorials* **26**(4), 2867–2913 (2024)
7. Berman, S.M., Greenhouse, D.S., Bailey, I.L., Clear, R.D., Raasch, T.W.: Human electroretinogram responses to video displays, fluorescent lighting, and other high frequency sources. *Optom. Vis. Sci.* **68**(8), 645–662 (1991)
8. Cahyadi, W.A., Chung, Y.H., Adiono, T.: Infrared indoor positioning using invisible beacon. In: 2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN), pp. 341–345 (2019)

9. Chen, F., Huang, N., Gong, C.: RSS-based visible light positioning with unknown receiver tilting angle: robust design and experimental demonstration. *Opt. Express* **30**(22), 39775–39793 (2022)
10. Cui, Z., Wang, Y., Fu, X.: Visible light positioning system based on RSSI fingerprint feature vector. In: 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), vol. 9, pp. 285–289 (2020)
11. Do, T.H., Yoo, M.: TDOA-based indoor positioning using visible light. *Photonic Netw. Commun.* **27**, 80–88 (2014)
12. Du, P., Zhang, S., Chen, C., Alphons, A., Zhong, W.D.: Demonstration of a low-complexity indoor visible light positioning system using an enhanced TDOA scheme. *IEEE Photonics J.* **10**(4), 1–10 (2018)
13. Farooq, M.S., Riaz, S., Abid, A., Abid, K., Naeem, M.A.: A survey on the role of iot in agriculture for the implementation of smart farming. *IEEE Access* **7**, 156237–156271 (2019)
14. Faulkner, N., Alam, F., Legg, M., Demidenko, S.: Watchers on the wall: passive visible light-based positioning and tracking with embedded light-sensors on the wall. *IEEE Trans. Instrum. Meas.* **69**(5), 2522–2532 (2020)
15. Hong, C.Y., et al.: Angle-of-arrival (AOA) visible light positioning (VLP) system using solar cells with third-order regression and ridge regression algorithms. *IEEE Photonics J.* **12**(3), 1–5 (2020)
16. Huang, T., et al.: Indoor 3D NLOS VLP using a single image sensor and two virtual LEDs. In: 2022 20th International Conference on Optical Communications and Networks (ICOON), pp. 1–3 (2022)
17. Huynh, P., Yoo, M.: VLC-based positioning system for an indoor environment using an image sensor and an accelerometer sensor. *Sensors* **16**(6) (2016). <https://www.mdpi.com/1424-8220/16/6/783>
18. Karanovic, O., Thabet, M., Wilson, H.R., Wilkinson, F.: Detection and discrimination of flicker contrast in migraine. *Cephalalgia* **31**(6), 723–736 (2011)
19. Keskin, M.F., Gezici, S.: Comparative theoretical analysis of distance estimation in visible light positioning systems. *J. Lightwave Technol.* **34**(3), 854–865 (2016)
20. Kim, H.S., Kim, D.R., Yang, S.H., Son, Y.H., Han, S.K.: An indoor visible light communication positioning system using a RF carrier allocation technique. *J. Lightwave Technol.* **31**(1), 134–144 (2013)
21. Kokdogan, F., Gezici, S.: Intelligent reflecting surfaces for visible light positioning based on received power measurements. *IEEE Trans. Veh. Technol.*, 1–14 (2024)
22. Li, Z., Yang, W., Xiao, L., Xiong, X., Wang, Z., Zou, X.: Integrated wearable indoor positioning system based on visible light positioning and inertial navigation using unscented kalman filter. In: 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1–6 (2019)
23. Luo, J., Fan, L., Li, H.: Indoor positioning systems based on visible light communication: state of the art. *IEEE Commun. Surv. Tutorials* **19**(4), 2871–2893 (2017)
24. Merenda, M., Catarinucci, L., Colella, R., Della Corte, F.G., Carotenuto, R.: Exploiting RFID technology for indoor positioning. In: 2021 6th International Conference on Smart and Sustainable Technologies (SpliTech), pp. 1–5 (2021)
25. Merenda, M., Catarinucci, L., Colella, R., Iero, D., Corte, F., Carotenuto, R.: RFID-based indoor positioning using edge machine learning. *IEEE J. Radio Freq. Ident.* **6**, 573–582 (2022)
26. Pan, Z., et al.: Design implementation of a hybrid VLP/PLC-based indoor tracking system for smart hospitals. In: 2023 IEEE International Conference on Consumer Electronics (ICCE), pp. 1–3 (2023)

27. Sisinni, E., Saifullah, A., Han, S., Jennehag, U., Gidlund, M.: Industrial internet of things: challenges, opportunities, and directions. *IEEE Trans. Industr. Inf.* **14**(11), 4724–4734 (2018)
28. Steendam, H.: A 3-D positioning algorithm for AOA-based VLP with an aperture-based receiver. *IEEE J. Sel. Areas Commun.* **36**(1), 23–33 (2018)
29. Steendam, H., Wang, T.Q., Armstrong, J.: Cramer-Rao bound for AOA-based VLP with an aperture-based receiver. In: 2017 IEEE International Conference on Communications (ICC), pp. 1–6 (2017)
30. Wang, K., Liu, Y., Hong, Z.: RSS-based visible light positioning based on channel state information. *Opt. Express* **30**(4), 5683–5699 (2022)
31. Wang, M., Ni, W., Shum, P.P., Yang, C.: High-precision indoor visible light positioning with tilt receiver based on image sensors. In: 2022 Asia Communications and Photonics Conference (ACP), pp. 546–550 (2022)
32. Wang, Q., et al.: A multimode fusion visible light localization algorithm using ambient lights. In: 2018 Ubiquitous Positioning, Indoor Navigation and Location-Based Services (UPINLBS), pp. 1–9 (2018)
33. Wang, T.Q., Sekercioglu, Y.A., Neild, A., Armstrong, J.: Position accuracy of time-of-arrival based ranging using visible light with application in indoor localization systems. *J. Lightwave Technol.* **31**(20), 3302–3308 (2013)
34. Wang, Z., et al.: POE-enabled visible light positioning network with low bandwidth requirement and high precision pulse reconstruction. *IEEE J. Indoor Seamless Positioning Navig.* **2**, 1–11 (2024)
35. Wen, S., et al.: Enhanced pedestrian navigation on smartphones with VLP-assisted PDR integration. *IEEE Sens. J.* **23**(14), 15952–15962 (2023)
36. Wilkins, A., et al.: A neurological basis for visual discomfort. *Brain* **107**(4), 989–1017 (1984). <https://doi.org/10.1093/brain/107.4.989>
37. Xie, C., Guan, W., Wu, Y., Fang, L., Cai, Y.: The LED-ID detection and recognition method based on visible light positioning using proximity method. *IEEE Photonics J.* **10**(2), 1–16 (2018)
38. Xu, S., Wu, Y., Wang, X., Wei, F.: Indoor high precision positioning system based on visible light communication and location fingerprinting. *J. Lightwave Technol.* **41**(17), 5564–5576 (2023)
39. Yang, S.H., Jung, E.M., Han, S.K.: Indoor location estimation based on LED visible light communication using multiple optical receivers. *IEEE Commun. Lett.* **17**(9), 1834–1837 (2013)
40. Yang, S.H., Kim, H.S., Son, Y.H., Han, S.K.: Three-dimensional visible light indoor localization using AOA and RSS with multiple optical receivers. *J. Lightwave Technol.* **32**(14), 2480–2485 (2014)
41. Yasir, M., Ho, S.W., Vellambi, B.N.: Indoor positioning system using visible light and accelerometer. *J. Lightwave Technol.* **32**(19), 3306–3316 (2014)



Evolving Trends in the Application of Artificial Intelligence to Support the Sustainable Development Goals: A Bibliometric Analysis

Kudakwashe Maguraushe¹ , Patrick Ndayizigamiye² , and Tebogo Bokaba³ 

¹ School of Computing, University of South Africa, Unisa Science Campus, Florida, Johannesburg, South Africa
magark@unisa.ac.za

² Centre for Applied Data Science, University of Johannesburg, Bunting Road Campus, Auckland Park, Johannesburg 2092, South Africa
pndayizigamiyep@uj.ac.za

³ Department of Applied Information Systems, University of Johannesburg, Bunting Road Campus, Auckland Park, Johannesburg 2092, South Africa
tbokaba@uj.ac.za

Abstract. This study conducted a bibliometric analysis to investigate the evolving trends in the application of artificial intelligence (AI) to support the Sustainable Development Goals (SDGs). A total of 2,653 peer-reviewed articles were retrieved from the Scopus database. The results indicate a substantial rise in publications between 2000 and 2024, and a particularly notable growth post-2018 reflecting increasing academic and practical interest. Noteworthy contributors, such as leading journals and influential papers, show the multidisciplinary nature of AI applications in addressing different SDGs. The geographical distribution of research shows that China and India are leading contributors, while the United States shows significant international collaboration. Key themes that arose include ‘machine learning’, ‘deep learning’, and ‘natural language processing’, with a significant focus on ‘sustainability’ and ‘societal concerns’. This study not only enriches comprehension of AI’s significance in supporting SDGs but also offers valuable insights for future research, policymaking, and practical application.

Keywords: Artificial Intelligence · Sustainable Development Goals · SDGs · Sustainable Development · Bibliometric Analysis · Interdisciplinary Research

1 Introduction

The Sustainable Development Goals (SDGs) established by the United Nations offer a comprehensive framework for tackling global challenges and promoting sustainable development [1]. In total, 17 SDGs were established in 2015 and these seek to tackle urgent global challenges such as poverty, inequality, climate change, environmental degradation, peace, and justice [2]. The SDGs are a set of global objectives that serve as a

comprehensive plan for attaining a more prosperous and environmentally friendly future for everyone [3]. Research endeavours and collaborations are essential in advancing progress towards the SDGs as societies work towards achieving the 2030 Agenda for Sustainable Development [4]. The integration of AI with SDGs gives a substantial prospect of applying advanced technology to achieve sustainable development. The discourse of the fusion of AI in the pursuit of the SDGs has gained prominence owing to its aptitude to improve processes, foster openness, and strengthen sustainable development endeavours [5].

This study sought to explore the evolving trends in the integration of AI in the quest to achieve the SDGs. It provides a summary of the present state of research and identifies prominent authors, institutions, countries, and emerging themes. Such a summary will aid in providing significant insights into the application of AI in addressing the SDGs and guide future research directions. Thus, this study had the following research objectives: 1. To explore the trends over time in the application of AI in supporting SDGs. 2. To identify key contributors and sources in the application of AI in supporting SDGs. 3. To identify the geographical distribution and collaboration in research related to the application of AI in the context of SDGs. 4. To identify core themes and influential works related to the application of AI in the context of SDGs. The significance of this study rests in its capacity to examine the scope of AI applications in sustainable development, pinpointing deficiencies, patterns, and prospective areas for collaboration. As AI technologies continue to evolve, it is becoming increasingly important to understand their role in attaining the SDGs. This analysis not only adds to the academic discussion but also serves as a valuable tool for policymakers, practitioners, and stakeholders who are dedicated to leveraging AI for sustainable development.

2 Background

SDGs acknowledge the inherent connection between people, the environment, and economic prosperity [6]. The SDGs consist of 17 goals with 169 relevant targets, which aim to tackle several global issues including poverty, inequality, climate change, environmental degradation, and many more [7–9]. The United Nations challenges all countries to attain the 17 SDGs goals by 2030 [10, 11]. Although they have become a focal point of development planning, the key challenge lies in determining how achievable the SDGs are given the present contexts [11]. Many critics have expressed concerns about the feasibility of these goals [10].

As AI is becoming regularly applied to accelerate progress towards the SDGs, it is, therefore, necessary to fully grasp the various uses of AI concerning the SDGs [1]. This will help understand the extent to which AI can help achieve the SDGs. For instance, AI technologies have been adopted to improve the quality of education (SDG 4) [5, 12], encourage actions that tackle climate change (SDG 13) [5, 13], promote the development of sustainable cities and communities (SDG 11) [14–16], and reinforce peace, justice, and robust institutions (SDG 16) [5, 13]. In addition, AI interventions have also demonstrated their potential to contribute to the provision of clean water and sanitation (SDG 6) [17, 18], and cheap and clean energy (SDG 7) [19, 20]. Through advanced data analytics, AI technologies support decision-making processes that are in line with SDG

9 (Industry, innovation, and infrastructure) [21, 22], and SDG 17 (Partnerships for the Goals) [21, 23]. Furthermore, AI has shown its potential to address issues leading to or related to poverty, and the stimulation of economic expansion in developing nations, highlighting its significance in attaining SDGs about poverty alleviation and infrastructure improvement [24]. The use of AI in the areas of food security, agriculture, and AgriTech operations exhibits its capacity to tackle issues related to food production and security [25]. This aligns with the SDGs that prioritize the establishment of sustainable food systems [25]. Thus, the potential of AI to contribute to the SDGs is extensive, as research suggests that AI has the potential to tackle as many as 134 out of the 169 targets stipulated by all 17 SDGs.

3 Methodology

3.1 Research Design

This study adopted bibliometric analysis to examine the evolving trends in the application of AI in supporting the attainment of the SDGs. Bibliometrics is a quantitative research method that enables the identification of patterns, trends, and key contributors in an identified field [26]. A quantitative research approach is used to methodically analyze a large dataset of peer-reviewed scholarly articles [27].

3.2 Database

The initial step in a bibliometric study is choosing the right database for document retrieval. This study utilized the Scopus database to retrieve peer-reviewed articles that were aligned with the study objectives. Scopus was chosen because it simplifies the process of acquiring large amounts of bibliometric data [28, 29]. In addition, exporting data from Scopus to other programs is also straightforward. Moreover, the Scopus search engine is more flexible as it provides basic and advanced search options for conducting precise searches effectively. Searching Scopus is achievable by using diverse terms connected to titles, abstracts, journals, authors, or affiliations. It enables analysis of the field's structure, dynamics, trends and evolution, and relationships among authors, articles, institutions, countries, and concepts.

3.3 Search Strategy

The following comprehensive search string was used to find relevant articles:

("artificial intelligence" OR "AI" OR "machine learning" OR "deep learning" OR "neural networks" OR "natural language processing" OR "computer vision" OR "reinforcement learning") AND ("sustainable development goals" OR "SDGs" OR "Millennium Development Goals" OR "no poverty" OR "zero hunger" OR "good health and well-being" OR "quality education" OR "gender equality" OR "clean water and sanitation" OR "affordable and clean energy" OR "decent work and economic growth" OR "industry innovation and infrastructure" OR "reduced inequalities" OR "sustainable cities and communities" OR "responsible consumption and production" OR "climate action")

OR “life below water” OR “life on land” OR “peace justice and strong institutions” OR “partnerships for the goals”). Data (peer-reviewed articles) was retrieved from the Scopus database on the 26th of July 2024 and spanned from 1992 to 2024. The retrieved peer-reviewed articles were analyzed using the Biblioshiny [30] and VOSviewer software [31]. Performance analysis (including publication trends, top contributing authors, top contributing journals, top contributing institutions, and top contributing countries) [26, 32], as well as science mapping (including relationships, co-occurrences, and network analysis) [26, 27], were conducted.

4 Results

4.1 Overview of the Research Landscape

The bibliometric data gathered from the Scopus database is summarized in Table 1. The table shows important indicators, including timespan, document information, author information, and document types. A total of 2,653 original articles were retrieved from the Scopus database. As depicted in Table 1, the retrieved articles were written by 8,561 authors, of which 283 were single authored. The table also shows a document co-authorship rate of 3.95, an average citation per document rate of 9.89, and a 31.89% international co-authorship rate, indicating a significant level of collaboration amongst authors from various countries. The significant number of publications in this field (2,653) indicates a strong and growing field of research. This is reflected in the large number of contributing authors, which shows broad academic interest. The significant number of citations per document demonstrates the influence and importance of these publications. Furthermore, the high rate of international collaboration shows the global scope of research efforts aimed at leveraging AI to support SDGs.

Table 1. Summary of articles on the application of AI in supporting SDGs.

Description	Result
Timespan	2002: 2024
Document information	Documents (2,653); Document average age (1.88); Average citations per doc (9.89); Annual growth (33.76%)
Author information	Authors involved (8,561); Authors of single-authored docs (283); Co-authors per document (3.95); International co-authorships (31.89%)
Document types	Journals; Book chapters; Books; Conference papers; Editorials; Reviews; Notes and letters

4.2 Trends Over Time in the Application of AI in Supporting SDGs

This section describes the results based on the growth and publication trends, highly cited papers, network analysis, the keywords that are mostly used, and trending topics in the

application of AI to support the SDGs. Figure 1 depicts the annual increase in scientific publications on the application of AI to the SDGs, from 2000 to 2024. It indicates a consistent rise in the number of publications over time, with a significant surge in recent years, specifically from 2018 to 2024. This could be attributed to the rapid developments in AI technologies and the heightened global focus on sustainable development.

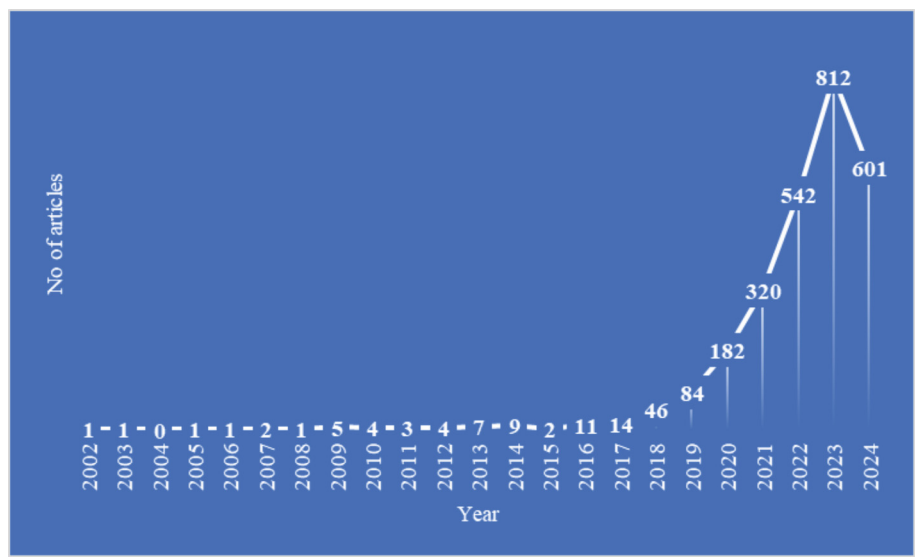


Fig. 1. Annual scientific production in the application of AI in supporting SDGs.

Figure 2 depicts the trend in the evolution-specific terms related to the intersection between AI and SDGs. The depicted trend offers valuable insights into the evolving research focus over time, highlighting emerging patterns and emerging areas of interest in this area. Furthermore, it highlights several emerging trends in AI research. There is increasing specialization within the field, as evidenced by the growing use of specific terms like “machine learning”, “deep learning”, and “natural language processing”.

4.3 Key Contributors and Sources

Figure 3 illustrates the evolution of contributions made by various academic journals throughout the years.

Figure 3 demonstrates that academic journals have undergone a rise in significance in disseminating research on AI and SDGs. This suggests the development of specialized publication outlets in this area. Figure 4 highlights the top 20 journals with the highest number of articles that portray integrated research on AI and SDGs. Prominent journals include *Sustainability* (127 articles), *Lecture Notes in Computer Science* (42), *Remote Sensing* (42), *Lecture Notes in Networks and Systems* (40), and *ACM International Conference* (31), which are leading sources for research on AI applications in supporting SDGs. The range of journals represented indicates the interdisciplinary nature of this

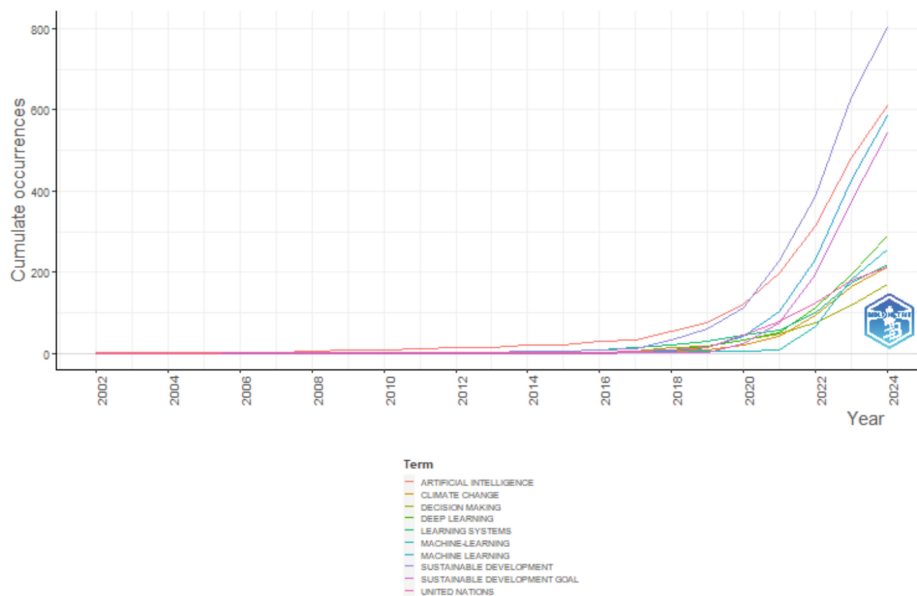


Fig. 2. Words’ frequency over time in the application of AI in supporting SDGs.

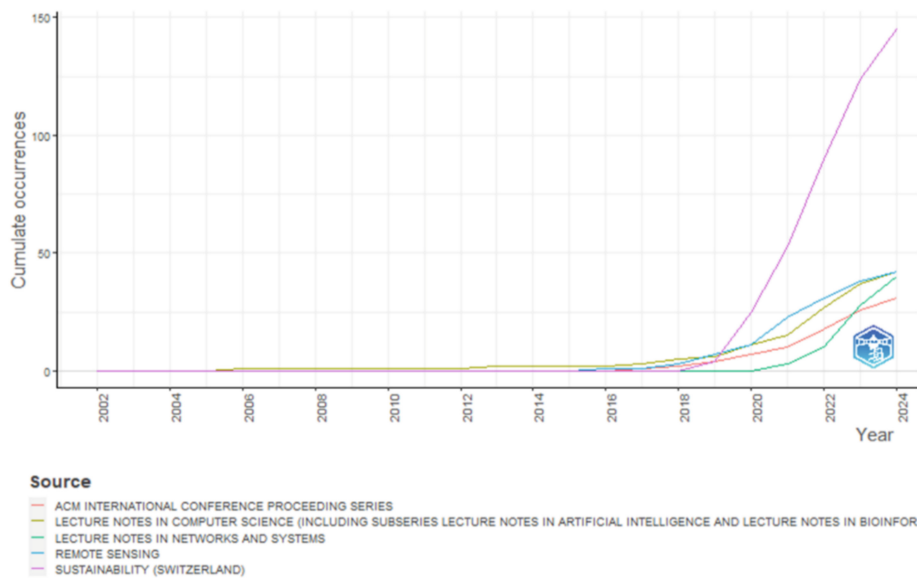


Fig. 3. Source production over time in the application of AI in supporting SDGs.

research area, with contributions from both AI-specific publications (such as *Artificial Intelligence Engineering*) and “broader environmental” or “social science” journals. This diversity reflects a widespread interest in AI’s role in addressing various SDGs.

Trends over time, as shown in Fig. 3, may indicate an increasing focus on this research area, with some journals expanding their output or dedicating special issues to AI and SDGs. Understanding these key journals aids researchers in choosing where to publish for optimal impact and helps emerging researchers in the field find relevant literature and stay current with research developments. Furthermore, recognizing leading journals can open opportunities for collaboration with editorial boards or researchers active in these outlets.

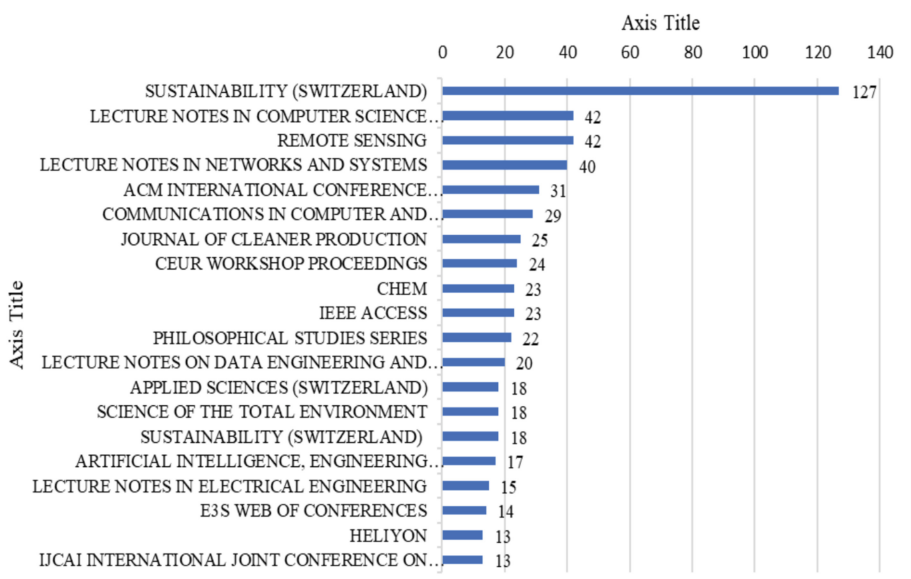


Fig. 4. Most relevant sources in the application of AI in supporting SDGs.

Further analysis reveals the top 20 journals using quantifiers such as the h-index, g-index, m-index, total citations (TC), and number of publications (NP), with a focus on identifying the most important sources. These are the most influential sources that have contributed to the use of AI in promoting the SDGs. The journal *Sustainability (Switzerland)* stands out as the most influential on the list, with 127 articles, an h-index of 26, a g-index of 42, an m-index of 4.33, and a total citation count of 2,273. These measures together demonstrate a significant rise in research on AI in the context of SDGs. Significantly, the journals *Remote Sensing*, *CHEM*, *Journal of Cleaner Production*, and *IEEE Access* are notable contributors to the knowledge base in this area. The information gives essential insights for researchers and policymakers who are navigating this dynamic research field.

Figure 5 depicts the number of articles produced by various institutions in the context of AI and SDGs. It demonstrates an expansive range of contributions from universities around the world. Uttarakhand University is at the forefront with 74 articles, while Beijing University closely follows with 63 articles. KTH Royal Institute of Technology and

Zhejiang University have contributed 38 and 36 articles, respectively. Significantly, several highly regarded institutions, such as the University of Johannesburg, King Fahd University of Petroleum and Minerals, and the University of California, have each produced 36, 34 and 33 articles, respectively.

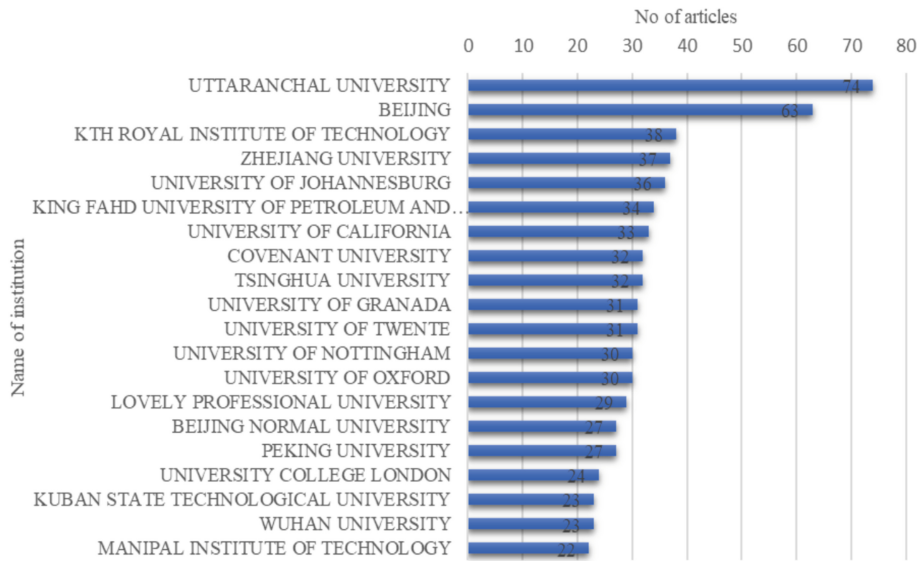


Fig. 5. Most relevant sources in the application of AI in supporting SDGs.

4.4 Geographical Distribution and Collaboration

Table 2 displays the geographical distribution of research contributions, highlighting the top 20 countries that are most actively publishing on the topic of AI and SDGs. Furthermore, it demonstrates the worldwide curiosity and collaborative efforts in this specific area of research. China is the top country in research that applies AI in the context of SDGs, with 283 articles. Many of these articles are single-country publications (SCP). India closely follows, demonstrating a significant research output. Although the United States has fewer publications, it exhibits a higher degree of collaboration, with 41 articles being accounted for by multiple country publications (MCP). South Africa also contributed 39 publications, indicating a varied geographical distribution of research on AI for SDGs, with countries from different continents actively involved in the research topic. The diverse landscape emphasizes the worldwide importance of AI in tackling sustainable development issues. The table also offers valuable information regarding the collaborative aspect of AI research for SDGs. The MCP ratio, which measures the percentage of papers involving international collaboration, differs throughout countries. While the United Kingdom and Canada have high rates of collaboration in their publications, China and India have a higher proportion of publications that are limited to a

single country. This indicates that there are varied levels of international collaboration in AI research for SDGs in the depicted countries.

Table 2. Top 20 Corresponding author countries in the application of AI in supporting SDGs.

Country	Articles	SCP	MCP	Freq	MCP_Ratio
China	283	199	84	0,1067	0,30
India	240	175	65	0,0905	0,27
USA	123	82	41	0,0464	0,33
United Kingdom	79	36	43	0,0298	0,54
Spain	67	42	25	0,0253	0,37
Australia	58	33	25	0,0219	0,43
Germany	56	43	13	0,0211	0,23
Italy	54	31	23	0,0204	0,43
South Africa	39	23	16	0,0147	0,41
Japan	38	29	9	0,0143	0,24
Canada	36	19	17	0,0136	0,47
Malaysia	36	17	19	0,0136	0,53
Saudi Arabia	36	14	22	0,0136	0,61
Sweden	34	16	18	0,0128	0,53
Nigeria	30	17	13	0,0113	0,43
Indonesia	28	20	8	0,0106	0,29
Iran	28	14	14	0,0106	0,50
Switzerland	25	15	10	0,0094	0,40
Brazil	24	19	5	0,0090	0,21
Netherlands	24	14	10	0,0090	0,42

TC - Total Citations; AAC - Average Article Citations; SCP - Single Country Publications; MCP - Multiple Country Publications

Figure 6 depicts the countries with the most frequently cited authors of studies on the application of AI for SDGs. The diagram presents a comparison between the total number of citations (TC) and the average number of citations per article for the top 20 countries in the field.

China has the highest number of cited authors, with the highest TC of 2,485 and an average citation per article of 8.8. The United States, with a TC of 2,423, demonstrates a higher average citation per article (19.7) than China. This suggests that the United States may have the highest influence per publication. Australia, Sweden, and the United Kingdom are in the next tier of influential countries. It is worth mentioning that although some countries, such as China, have a significant number of overall citations, their average citation per article may be comparatively lower. This may indicate a greater

quantity of articles, but potentially a lower impact per individual article. In contrast, countries like Sweden (48.4), Singapore (39.1), Australia (29.4), Netherlands (25), and Switzerland (22.5) have higher average citations per article. Although these countries may have fewer total publications, every publication they produce tends to have a greater impact based on their citations. In essence, Fig. 6 provides useful insights into the worldwide AI for SDGs research landscape, emphasizing the top countries and their corresponding research influence (based on their citations). These findings can provide valuable insights for international partnerships sharing of knowledge, and impact of research per country.

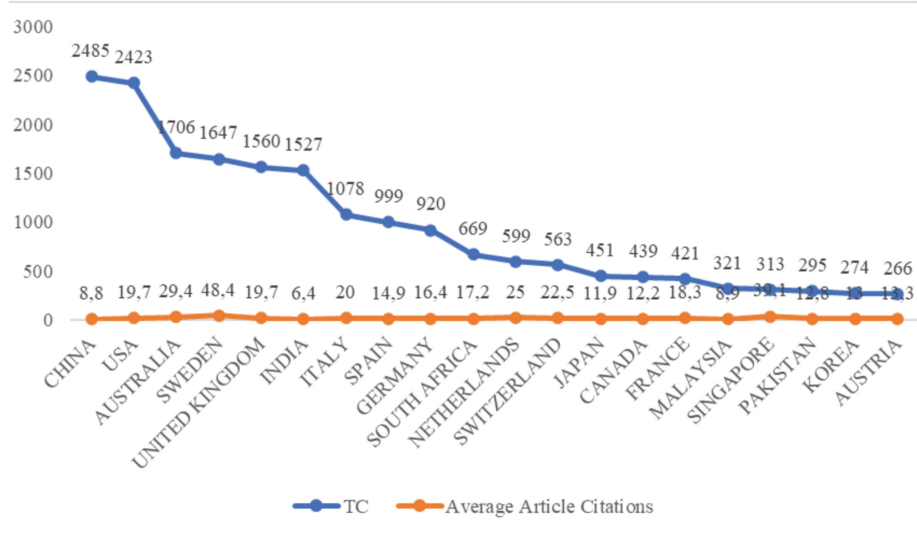


Fig. 6. Most cited countries in the application of AI in supporting SDGs.

4.5 Core Themes and Influential Works

Figure 7 displays a Word Cloud that shows how often certain keywords connected to the use of AI in the context of the SDGs occur. The size of each word is directly proportional to its frequency throughout the dataset, that is the entire set of retrieved papers. The Word Cloud highlights multiple key themes. The presence of terms such as “artificial intelligence,” “machine learning,” and “deep learning” is prominent, implying a significant emphasis on AI technology in the dataset. Similarly, terms such as “sustainable development,” “SDGs,” and “climate change” emphasize the ecological and societal aspects of research conducted in the context of AI and SDGs. The terms “data mining,” “big data,” and “prediction” highlight the significance of utilizing data-driven methods in the application of AI for achieving SDGs. The terms “Internet of Things,” “neural networks,” and “algorithms” indicate the technological foundations of the research. The keywords “agriculture,” “forestry,” “water quality,” and “food supply” indicate a wide

variety of sectors where AI can be applied to tackle SDGs. The inclusion of terms such as “humans,” “education,” and “decision-making” suggests a recognition of the impact of AI on human beings and the broader implications for SDGs.

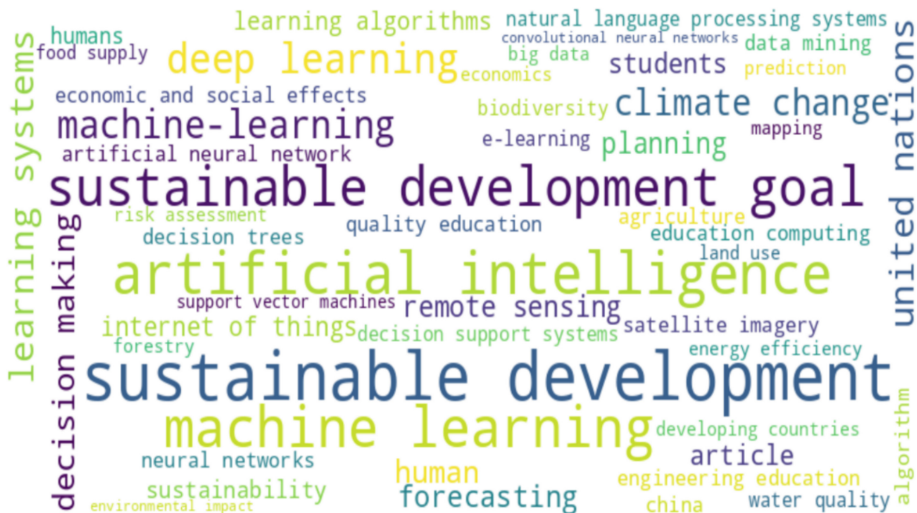


Fig. 7. Word Cloud in the application of AI in supporting SDGs.

Figure 8 displays a dynamic visualization that shows how research interests in the subject of AI for SDGs have changed over time. The diagram uses a chronological structure, monitoring the occurrence rate of terms from 2010 to 2024.

Each term is depicted by a horizontal line with a diamond shape, where the size of the shape is proportional to the frequency of the term at a specific moment in time. The continual prevalence and expansion of terms such as “artificial intelligence,” “machine learning,” “deep learning,” and “semantic segmentation” highlight the pivotal significance of AI technologies in tackling SDGs. Concepts like “sustainable development,” “SDGs,” and “climate change” consistently remain in the spotlight, highlighting the continued emphasis on environmental and social sustainability. The rising significance of phrases such as “big data,” “data mining,” and “information technology” indicates the increasing focus on data-driven methodologies and analytics for making well-informed decisions. The information provided in Fig. 8 is crucial for spotting new trends, guiding research priorities, and promoting interdisciplinary collaborations.

Further analysis revealed the top 20 most cited papers on AI and SDGs. Leading authors include Vinuesa et al. (2020) whose article “The role of artificial intelligence in achieving the Sustainable Development Goals” published in *Nature Communications* has 1,069 citations, averaging 213.80 citations per year. Allam & Dhunny (2019) explored AI and smart cities, receiving 657 citations and an average of 109.50 per year. Di Vaio et al. (2020) reviewed AI and business models for SDGs in the *Journal of Business Research*, with 471 citations and an annual average of 94.20. Mondejar et al. (2021) examined digitalization for SDGs in the journal *Science of the Total Environment*, earning 343

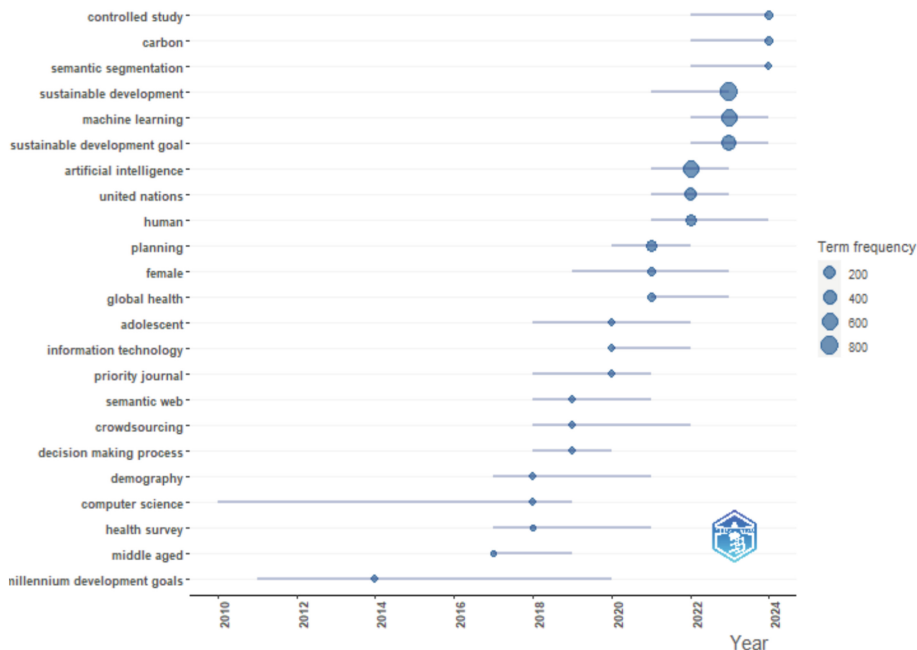


Fig. 8. Trending terms in the application of AI in supporting SDGs.

citations and 85.75 per year. Parkatzidis et al. (2020) provided an update on controlled radical polymerization in *Chem*, achieving 318 citations at a rate of 63.60 per year.

Figure 9 is a thematic map that depicts the linkages between different themes associated with AI and SDGs. The themes are arranged according to their “development degree” (density) and “relevance degree” (centrality). The motor themes are situated in the upper right quadrant, indicating their central position and significant level of development. The themes encompassed are “climate change”, “sustainability”, “algorithm”, “sustainable development”, “artificial intelligence” and “sustainable development goals”. The niche themes are in the upper left quadrant, and although they are less central, they nevertheless exhibit notable progress. Their primary focus is on matters associated with humanity. The basic themes are situated in the lower right quadrant, and these hold a central position but display little development. The themes encompassed are “machine learning”, “learning systems” and “deep learning”. Finally, emerging, or declining themes are situated in the lower left quadrant, characterized by their low centrality, and limited development. The themes encompassed are “quality education”, “students”, “education computing”, “natural language processing systems”, “natural language processing”, and “social media”. The map encompasses a group of interconnected themes centred around “sustainable development”, “artificial intelligence”, and “machine learning”, indicating a significant linkage across these domains. There is a notable emphasis on human-centred topics as evidenced by the frequent use of terms such as “human” and “humans”. Moreover, the inclusion of themes relating to “education” suggests an inclination towards harnessing AI for educational purposes.

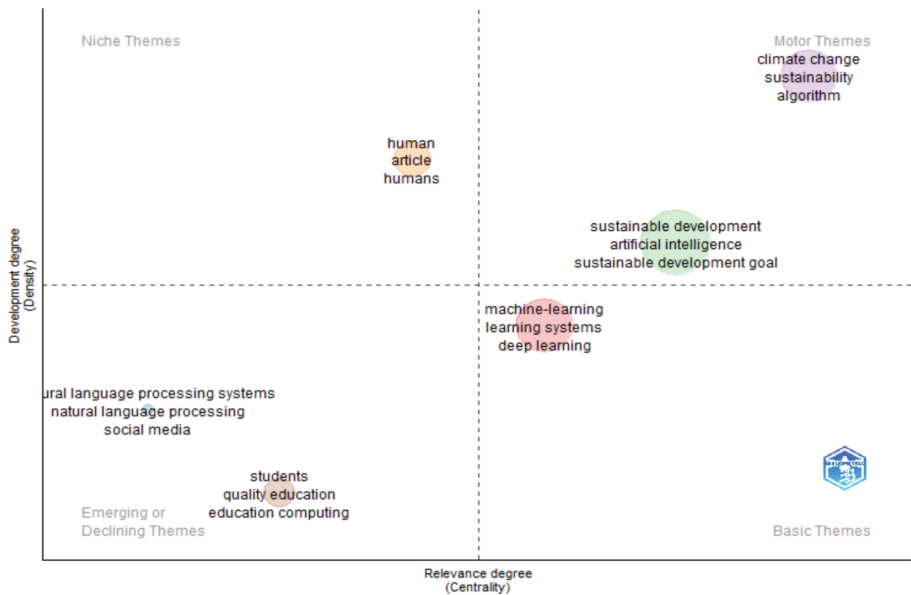


Fig. 9. Thematic map in the application of AI in supporting SDGs.

5 Discussion

This bibliometric research offers an extensive summary of the evolving trends in the application of AI to facilitate the achievement of the SDGs. This study provides insights into the current state of research in this interdisciplinary field by analyzing publishing patterns, citation trends, and emerging themes. It also identifies important contributors, geographical distributions, collaboration networks, and core themes. The review of bibliometric data from the Scopus database demonstrates an influential and increasing amount of research in the field of AI applications that contribute to the advancement of SDGs. Flagging the “bandwagon effect” in this field, [33] demonstrated that Scopus-indexed outputs reflect significant SDG researchers, publication outlets, and collaboration patterns. The field exhibits significant academic interest and activity, as evidenced by a total of 2,653 original articles produced by 8,561 researchers. The mean citation rate of 9.89 and the international co-authorship rate of 31.89% highlight the importance and global scope of research endeavors in this field. The significant degree of international collaboration shows the significance of cross-border collaborations in tackling complex global concerns using AI. A study [34] concurred with this submission and argued that this phenomenon is a component of a wider surge in research related to SDGs, suggesting a rising consciousness of global issues and a transition towards interdisciplinary strategies.

The study demonstrated a steady increase in publications from 2000 to 2024, with a significant surge in more recent years (2018–2024). The rise in interest in utilizing AI for sustainable development is a result of increased academic and practical attention. As posited by [35], the period between 2015 and 2022 witnessed a significant increase

in AI research specifically focused on health-related SDGs. This is likely due to the quick progress in technology and the heightened worldwide emphasis on sustainability. The changing research emphasis is seen in the prevalence of certain words like “machine learning,” “deep learning,” and “natural language processing,” suggesting a transition towards specialized applications of AI. Concurrently, the focus on concepts such as “sustainability,” “water,” “agriculture,” “biodiversity,” and “quality of education” emphasizes how AI research is in line with specific SDGs. This is in tandem with [5] who demonstrated that AI applications in sustainable development encompass prediction algorithms, machine learning techniques, and neural networks. This helps shape policy and financing decisions and brings attention to future research areas.

A look at contributions from different academic journals demonstrates the growing importance of specialized publications in distributing research on AI and SDGs. Renowned publications including *Sustainability (Switzerland)*, *Lecture Notes in Computer Science*, and *Remote Sensing* have established themselves as primary references, showcasing the interdisciplinary character of the research, which was also confirmed by [36]. The journals, including *Sustainability (Switzerland)*, have exceptional effect and influence, as evidenced by their h-index, g-index, m-index, and total citation counts. The broad spectrum of journals reflects the extensive interest in the role of AI in tackling different SDGs, offering valuable insights for academics who are looking for the best outlets to publish their work and opportunities for collaboration. As alluded to by [37], remote sensing research predominantly centres on SDG 15 (life on land) and 11 (sustainable cities), with China and the United States taking the lead in collaborations.

Geographical analysis reveals a varied distribution of research contributions, with China leading in terms of the number of articles, closely followed by India and the United States. Specifically, since 2010, China has become an important factor in global collaborations related to AI [38]. The significant levels of international collaboration, specifically from the United States and the United Kingdom, highlight the worldwide impact of AI in tackling SDGs and the significance of collaborations between different countries.

The thematic analysis reveals several key themes, including “artificial intelligence,” “machine learning,” “deep learning,” “sustainable development,” “climate change,” and “big data.” These themes emphasize the technological, ecological, and societal aspects of AI research about SDGs. The application of AI technology in sustainable development is highlighted by the Word Cloud and dynamic depiction of research interests throughout time, a phenomenon which was widely confirmed by [5]. The thematic map also provides additional visual evidence of the relationship between various themes, highlighting the notable progress in motor-related themes such as “climate change,” “sustainability,” and “algorithm.” Emerging themes like “natural language processing systems” and “social media” suggest areas that are gaining attention and could be viable routes for future research.

6 Implications of the Study

The findings of this bibliometric analysis offer valuable insights for academics, policymakers, and practitioners. The growing body of research underscores the need for continued investment in AI for sustainable development. Policymakers should prioritize

funding and supporting cross-disciplinary research to address complex global challenges through AI. Major contributors and collaboration networks identified in this study highlight the importance of fostering partnerships and knowledge exchange. For researchers, engaging in interdisciplinary and international collaborations can broaden the impact and reach of their work. Emerging themes and influential studies suggest promising avenues for future exploration, including the ethical dimensions of AI, the scalability of AI solutions in low-resource environments, and integration with other technologies like IoT and blockchain. These areas hold potential for advancing AI's role in achieving sustainable development goals on a global scale.

6.1 Theoretical Implications

Theoretical insights reveal a need for frameworks that support responsible AI design, particularly in ethical and diverse contexts. Current models do not fully address the social complexities AI encounters when applied to sustainable development, suggesting room for interdisciplinary theories that incorporate ethical, cultural, and environmental factors. These theories could help clarify AI's potential and limitations within the SDGs, shaping guidelines for responsible and context-aware AI applications that align with long-term sustainable development goals.

6.2 Practical Implications

AI's practical contributions to the SDGs are already visible in areas like climate action, healthcare, and education. Predictive models driven by AI are helping communities prepare for extreme weather events, while healthcare applications are enhancing diagnostics and treatment accessibility, especially in underserved areas. In education, AI supports personalized learning, making it easier to meet the needs of diverse students. However, ethical challenges persist, particularly around data privacy and algorithmic bias, underscoring the need for responsible AI practices that prioritize transparency, inclusivity, and sustainability.

6.3 Limitations

This study, while informative, has a few important limitations. First, it relies solely on data from Scopus, which may overlook relevant research in other databases, potentially limiting the scope. Additionally, the study is largely quantitative, which captures broad trends but misses deeper qualitative insights from practitioners and case studies. Another limitation is the timeframe, which may miss the most recent shifts in this fast-evolving field. Lastly, the breadth of SDG topics means findings may not apply equally across all goals. AI's role in some areas, like poverty reduction, is less represented and could benefit from dedicated research attention.

6.4 Future Research

Future research in AI for sustainable development can take several meaningful directions. A key focus could be on creating AI models that are tailored to the unique challenges and

resources of different regions, enhancing local relevance and effectiveness. Ethical considerations are also crucial and future research could establish frameworks that prioritize privacy and data ownership, which would help build trust among users. Another vital area is examining the long-term societal impact of AI on things like workforce dynamics and community resilience, as the effects of these technologies unfold over time. There is also a growing need for “Green AI” solutions that reduce energy demands, aligning with environmental SDGs. Equally important, interdisciplinary collaborations can provide a well-rounded approach to AI in sustainable development, bringing together insights from various fields to address complex, multi-dimensional goals.

7 Conclusion

This study emphasized the growing impact of AI in advancing the SDGs, with research showing a strong global and interdisciplinary effort. Key players and emerging areas of focus point to exciting uses of AI, from enhancing climate resilience to improving healthcare access, though ethical issues like data privacy and fairness are essential to address. To fully harness AI’s potential for sustainable development, future efforts should prioritize responsible, adaptable, and collaborative approaches. This analysis provides a valuable foundation for researchers, policymakers, and practitioners aiming to align AI innovations with real-world sustainability goals.

References

1. Vinuesa, R., et al.: The role of artificial intelligence in achieving the Sustainable Development Goals. *Nat. Commun.* **11**(1), 1 (2020). <https://doi.org/10.1038/s41467-019-14108-y>
2. U. Nations. ‘The 17 Goals’ (2024). <https://sdgs.un.org/goals>. Accessed 4 Aug 2024
3. Triatmanto, B., Natsir, M.: A Descriptive Analysis of Sustainability Development Goals (SDGs), community empowerment, and the participatory rural appraisal approach’. In: 6th International Conference on Community Development, pp. 118–121 (2019). <https://doi.org/10.2991/iccd-19.2019.32>
4. Sweileh, W.M.: Bibliometric analysis of scientific publications on “sustainable development goals” with emphasis on “good health and well-being” goal (2015–2019). *Global. Health* **16**(1), 1–13 (2020). <https://doi.org/10.1186/s12992-020-00602-2>
5. Singh, A., Kanaujia, A., Singh, V.K., Vinuesa, R.: Artificial intelligence for Sustainable Development Goals: Bibliometric patterns and concept evolution trajectories. *Sustain. Dev.* **32**(1), 724–754 (2024). <https://doi.org/10.1002/sd.2706>
6. Donaires, O.S., Cezarino, L.O., Caldana, A.C.F., Liboni, L.: Sustainable development goals – an analysis of outcomes. *Kybernetes* **48**(1), 183–207 (2019). <https://doi.org/10.1108/K-10-2017-0401>
7. Halkos, G., Gkampoura, E.C.: Where do we stand on the 17 Sustainable Development Goals? An overview on progress. *Econ. Anal. Policy* **70**, 94–122 (2021). <https://doi.org/10.1016/j.eap.2021.02.001>
8. Fuso Nerini, F., et al., ‘Mapping synergies and trade-offs between energy and the Sustainable Development Goals’. *Nat. Energy* **3**(1), 10–15 (2018). <https://doi.org/10.1038/s41560-017-0036-5>
9. Carlsen, L., Bruggemann, R.: The 17 United Nations’ sustainable development goals: a status by 2020. *Int. J. Sustain. Dev. World Ecol.* **29**(3), 219–229 (2022). <https://doi.org/10.1080/13504509.2021.1948456>

10. Bali Swain, R., Yang-Wallentin, F.: Achieving sustainable development goals: predicaments and strategies. *Int. J. Sustain. Dev. World Ecol.* **27**(2), 96–106 (2020). <https://doi.org/10.1080/13504509.2019.1692316>
11. Moyer, J.D., Hedden, S.: Are we on the right path to achieve the sustainable development goals? *World Dev.* **127**, 104749 (2020). <https://doi.org/10.1016/j.worlddev.2019.104749>
12. Okulich-Kazarin, V., Artyukhov, A., Skowron, Ł., Artyukhova, N., Wołowicz, T.: Will AI become a threat to higher education sustainability? a study of students' Views. *Sustain.* **16**(11), 1–19 (2024). <https://doi.org/10.3390/su16114596>
13. Wakunuma, K., Ogoh, G., Eke, D.O., Akintoye, S.: Responsible AI, SDGs, and AI governance in Africa'. In: 2022 IST-Africa Conference, IST-Africa 2022, IST-Africa Institute and Authors, pp. 1–13 (2022). <https://doi.org/10.23919/IST-Africa56635.2022.9845598>
14. Zaidi, A., Ajibade, S.S.M., Musa, M., Bekun, F.V.: New insights into the research landscape on the application of artificial intelligence in sustainable smart cities: a bibliometric mapping and network analysis approach. *Int. J. Energy Econ. Policy* **13**(4), 287–299 (2023). <https://doi.org/10.32479/ijeep.14683>
15. Gupta, S., Degbelo, A.: An empirical analysis of AI contributions to sustainable cities (SDG 11). *Philos. Stud. Ser.* **152**, 461–484 (2023). https://doi.org/10.1007/978-3-031-21147-8_25
16. Si, D.: A framework to analyze the Impacts of AI with the sustainable development goals. *Highlights Sci. Eng. Technol.* **17**, 313–323 (2022). <https://doi.org/10.54097/hset.v17i.2621>
17. Dada, M.A., Majemite, M.T., Obaigbena, A., Daraojimba, O.H., Oliha, J.S., Nwokediegwu, Z.Q.S.: Review of smart water management: IoT and AI in water and wastewater treatment. *World J. Adv. Res. Rev.* **21**(1), 1373–1382 (2024). <https://doi.org/10.30574/wjarr.2024.21.1.0171>
18. Das, S.C., Kumar Sanyal, M., Das2, J.K., Bhattacharyya, N.: Clean Water, Sanitation and zero emission Industry Pollution Management using nWSN based Green IoT Applications: a Techno-legal impact. *Int. J. Adv. Eng. Manag.* **5**(03), 605 (2023). www.ijaem.net
19. Chaudhary, G.: Environmental sustainability: can artificial intelligence be an enabler for SDGs? *Nat. Environ. Pollut. Technol.* **22**(3), 1411–1420 (2023). <https://doi.org/10.46488/NEPT.2023.v22i03.027>
20. Yadav, M., Singh, G.: Environmental sustainability with artificial intelligence EPRA Int. J. Multidiscip. Res. (IJMR)-Peer Rev. J., **2**, 198–210 (2020). <https://doi.org/10.36713/epra2013>
21. Kulkov, I., Kulkova, J., Rohrbeck, R., Menvielle, L., Kaartemo, V., Makkonen, H.: Artificial intelligence - driven sustainable development: Examining organizational, technical, and processing approaches to achieving global goals. *Sustain. Dev.* **32**(3), 2253–2267 (2024). <https://doi.org/10.1002/sd.2773>
22. Sadic, S., Demir, E., Crispim, J.: Towards a connected world: Collaborative networks as a tool to accomplish the SDGs. *J. Clean. Prod.* **462**, (2024). <https://doi.org/10.1016/j.jclepro.2024.142726>
23. Yu, Z., Liang, Z., Wu, P.: How data shape actor relations in artificial intelligence innovation systems: an empirical observation from China. *Ind. Corp. Chang.* **30**(1), 251–267 (2021). <https://doi.org/10.1093/icc/dtaa063>
24. Mhlanga, D.: Artificial intelligence in the industry 4.0, and its impact on poverty, innovation, infrastructure development, and the sustainable development goals: lessons from emerging economies?. *Sustain.* **13**(11) (2021). <https://doi.org/10.3390/su13115788>
25. Spanaki, K., Karafili, E., Sivarajah, U., Despoudi, S., Irani, Z.: Artificial intelligence and food security: swarm intelligence of AgriTech drones for smart AgriFood operations. *Prod. Plan. Control* **33**(16), 1498–1516 (2022). <https://doi.org/10.1080/09537287.2021.1882688>
26. Passas, I.: Bibliometric analysis: the main steps. *Encyclopedia* **4**, 1014–1025 (2024)
27. Öztürk, O., Kocaman, R., Kanbach, D.K.: How to design bibliometric research: an overview and a framework proposal. *Rev. Manag. Sci.* 0123456789 (2024). <https://doi.org/10.1007/s11846-024-00738-0>

28. Sweileh, W.M.: Bibliometric analysis of peer-reviewed literature on climate change and human health with an emphasis on infectious diseases. *Global. Health* **16**(1), 1–17 (2020). <https://doi.org/10.1186/s12992-020-00576-1>
29. Maguraushe, K., Ndayizigamiye, P., Bokaba, T.: Trends and developments in the use of machine learning for disaster management: a bibliometric analysis, vol. 698 AICT. Springer Nature Switzerland (2024). https://doi.org/10.1007/978-3-031-50192-0_9
30. Fu, Y., Mao, Y., Jiang, S., Luo, S., Chen, X., Xiao, W.: A bibliometric analysis of systematic reviews and meta-analyses in ophthalmology. *Front. Med.* **10**(March), 1–9 (2023). <https://doi.org/10.3389/fmed.2023.1135592>
31. Mejia, C., Wu, M., Zhang, Y., Kajikawa, Y.: Exploring topics in bibliometric research through citation networks and semantic analysis. *Front. Res. Metrics Anal.* **6**(September), 1–16 (2021). <https://doi.org/10.3389/frma.2021.742311>
32. Lim, W.M., Kumar, S.: Guidelines for interpreting the results of bibliometric analysis: a sensemaking approach. *Glob. Bus. Organ. Excell.* **43**(2), 17–26 (2024). <https://doi.org/10.1002/joe.22229>
33. Alfrević, N., Malešević Perović, L., Mihaljević Kosor, M.: Productivity and Impact of Sustainable Development Goals (SDGs)-related academic research: a bibliometric analysis. *Sustain.* **15**(9) (2023). <https://doi.org/10.3390/su15097434>
34. Yumnam, G., Gyanendra, Y., Singh, C.I.: A systematic bibliometric review of the global research dynamics of United Nations Sustainable Development Goals 2030. *Sustain. Futur.* **7**, no. March, 100192 (2024). <https://doi.org/10.1016/j.sfr.2024.100192>
35. Ramezani, M., Takian, A., Bakhtiari, A., Rabiee, H.R., Sazgarnejad, S.: Bibliometric analysis of artificial intelligence revolutions in health-related sustainable development goals. *Heal. Technol. Assess. Action* **7**(4) (2023). <https://doi.org/10.18502/htaa.v7i4.14654>
36. Meschede, C.: The sustainable development goals in scientific literature: a bibliometric overview at the meta-level. *Sustain.* **12**(11) (2020). <https://doi.org/10.3390/su12114461>
37. Ekmen, O., Kocaman, S.: Remote sensing for UN SDGs: a global analysis of research and collaborations. *Egypt. J. Remote Sens. Sp. Sci.* **27**(2), 329–341 (2024). <https://doi.org/10.1016/j.ejrs.2024.04.002>
38. Tang, X., Li, X., Ding, Y., Ma, F.: Understanding team collaboration in artificial intelligence from the perspective of geographic distance', *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence Lecture Notes Bioinformatics)*, vol. 12645 LNCS, pp. 14–23 (2021). https://doi.org/10.1007/978-3-030-71292-1_2

Author Index

A

Abba Ari, Ado Adamou II-144
Abisoye, Blessing Olatunde II-291
Aboagye, Manhyia Gideon II-275
Aboagye, Sylvester I-240
Agyemang, Justice Owusu I-152, II-275
Ahiaklo-Kuz, Raphael I-240
Akoguhi, N'guessan Patrice I-210
Aliyu, Ahmed II-194
Alsolame, Tarek I-193
Amaury, Cansa Kayembe II-177
Antogiovanni, Stephen I-193
Aritra, B. D. S. I-193
Assohoun, Egomli Stanislas I-210
Atakpla, Desmond Delali I-152

B

Ba, Cheikh II-105
Bagha, Amir Mohammadi II-164
Bagula, Antoine I-83, II-260
Boadu, Thomas I-152
Bokaba, Tebogo I-275, II-70, II-194
Bueno, Felipe I-252
Byenkyia, Nkalubo Lenard II-17, II-125

C

Chakrabortii, Chandranil I-193
Cheng, Ling I-228
Clinton, Addai II-275

D

Darko, Christabel Bemah I-152
Diakhame, Mouhamadou Lamine II-213
Diattara, Awa II-105
Dieng, Diarra Bousso II-213
Dogo, Eustace M. II-194
Dombeu, Jean Vincent Fonou II-36, II-87
Dtissibe, Francis Yongwa II-144
Du Plessis, Deon I-133

E

Elikplim, Henyo Chester II-275

G

Gueroui, Abdelhak Mourad II-144
Gwetu, Mandlenkosi V. I-97
Gwetu, Mandlenkosi Victor II-36
Gwetu, Mandlenkosi II-55, II-87, II-227

H

Halefose, Kholofelo H. I-133
Hassanalizadeh, Azin II-164
Henehan, Thomas II-369

I

Igboamalu, Frank I-228

J

Jonathan, Kabemba Ntumbwa II-177

K

Kandji, Abdou Karim II-369
Kodua-Basoa, Martin I-152
Kogeda, Okuthe P. I-34
Koua, Kpangni Alex Jérémie I-210, II-326
Kouame, Koffi Fernand I-210
Kouame, N'dri Pascal I-210
Kparbong, Ponjul Bali II-194
Kwenda, Clopas II-36
Kyangwa, Henriette I-83

L

Leke, Collins A. I-3
Lekola, Bafokeng II-347
Lekola, Serestina II-347
Lerato, Moila Ramahlapane I-177
Lumamba, Kapongo D. I-97

M

Maguraushe, Kudakwashe I-275
 Makun, Hussaini A. II-194
 Mathonsi, Topside E. I-133
 Mbale, Landry I-83, II-260
 Merlec, Mpyana Mwamba II-177
 Mhlongo, Siyabonga II-70
 Mienye, Ibomoie Domor II-3
 Mngomezulu, Mangaliso II-55
 Molefe, Mohale II-309
 Moodley, Larisha I-34
 Msimnag, Mpumelelo I-97
 Mthulisi, Velepini I-177
 Muchenje, Tonderai I-133
 Mwaura, Jonathan I-121, II-213, II-326,
 II-369

N

Naidoo, Threnesan I-97
 Ndayizigamiye, Patrick I-275
 Ndjiongue, Alain R. I-3
 Ndjiongue, Alain I-228
 Noundam, René II-144
 Ngandu, Bernard II-260
 Ngossaha, Justin Moskolai II-144
 Nixson, Okila II-17, II-125
 Niyonkuru, Enock I-193
 Nougaret, Joshua II-213
 Nyasulu, Chimango I-121

O

Obaido, George II-3
 Obed, Somuah II-275
 Ofusori, Lizzy II-70
 Olawuyi, Olushola I-15, I-54
 Ouahada, Khmaies I-228
 Owolawi, Pius I-166

P

Pouamoun, Ibrahim Aoudou II-144

R

Rawat, Danda B. II-164
 Rose, Nakibuule II-17, II-125
 Rotshidzwa, Phandavhudzi I-177
 Rundensteiner, Elke II-105

S

Saghir, Zohaib I-240
 Sané, Fatou Cire Emilie II-326
 Selorm, Awumee Gabriel I-152, II-275
 Seretloa, Semakaleng I-166
 Sibanda, Khulumani I-72
 Sibitenda, Harriet II-105
 Singh, Yuvika I-34
 Steyn, Adrie J. C. I-97
 Sun, Yanxia II-291
 Swart, Theo G. II-3

T

Tamuka, Nyashadzashe I-72
 Tanwar, Sudeep II-164
 Tchaleu, Boriane Y. I-3
 Tivir, Cecilia I-121
 Traore, Assitan II-105
 Traore, Issa II-164
 Tu, Chunling I-166

V

Vambe, Trevor II-227
 Viriri, Serestina II-227

W

Warren, Michael II-326
 Wells, Gordon I-97
 Woungang, Isaac II-164

Z

Zengeya, Tsitsi II-87
 Zenghui, Wang II-291
 Zinsou, Kpêtchéhoué Merveille Santi II-369