Pertanika Journal of

SCIENCE & TECHNOLOGY

JST

VOL. 33 (6) OCT. 2025

PERTANIKA JOURNALS

A scientific journal published by Universiti Putra Malaysia Press

Pertanika Journal of Science & Technology
Pertanika Journal of Science & Technology

PERTANIKA JOURNAL OF SCIENCE & TECHNOLOGY

About the Journal

Overview

Pertanika Journal of Science & Technology is an official journal of Universiti Putra Malaysia. It is an open-access online scientific journal. It publishes original scientific outputs. It neither accepts nor commissions third party content.

Recognised internationally as the leading peer-reviewed interdisciplinary journal devoted to the publication of original papers, it serves as a forum for practical approaches to improve quality on issues pertaining to science and engineering and its related fields.

Pertanika Journal of Science & Technology currently publishes 6 issues a year (*January, March, April*, *July, August,* and *October*). It is considered for publication of original articles as per its scope. The journal publishes in **English** and it is open for submission by authors from all over the world.

The journal is available world-wide.

Aims and scope

Pertanika Journal of Science & Technology aims to provide a forum for high quality research related to science and engineering research. Areas relevant to the scope of the journal include: bioinformatics, bioscience, biotechnology and bio-molecular sciences, chemistry, computer science, ecology, engineering, engineering design, environmental control and management, mathematics and statistics, medicine and health sciences, nanotechnology, physics, safety and emergency management, and related fields of study.

History

Pertanika Journal of Science & Technology was founded in 1993 and focuses on research in science and engineering and its related fields.

Vision

To publish a journal of international repute.

Mission

Our goal is to bring the highest quality research to the widest possible audience.

Quality

We aim for excellence, sustained by a responsible and professional approach to journal publishing. Submissions can expect to receive a decision within 90 days. The elapsed time from submission to publication for the articles averages 180 days. We are working towards decreasing the processing time with the help of our editors and the reviewers.

Abstracting and indexing of Pertanika

Pertanika Journal of Science & Technology is now over 33 years old; this accumulated knowledge and experience has resulted the journal being abstracted and indexed in SCOPUS (Elsevier), Journal Citation Reports (JCR-Clarivate), EBSCO, ASEAN CITATION INDEX, Microsoft Academic, Google Scholar, and MyCite.

Citing journal articles

The abbreviation for Pertanika Journal of Science & Technology is *Pertanika J. Sci. & Technol.*

Publication policy

Pertanika policy prohibits an author from submitting the same manuscript for concurrent consideration by two or more publications. It prohibits as well publication of any manuscript that has already been published either in whole or substantial part elsewhere. It also does not permit publication of manuscript that has been published in full in proceedings.

Code of Ethics

The *Pertanika* journals and Universiti Putra Malaysia take seriously the responsibility of all of its journal publications to reflect the highest in publication ethics. Thus, all journals and journal editors are expected to abide by the journal's codes of ethics. Refer to *Pertanika*'s **Code of Ethics** for full details, or visit the journal's web link at http://www.pertanika.upm.edu.my/code_of_ethics.php

Originality

The author must ensure that when a manuscript is submitted to *Pertanika*, the manuscript must be an original work. The author should check the manuscript for any possible plagiarism using any program such as Turn-It-In or any other software before submitting the manuscripts to the *Pertanika* Editorial Office, Journal Division.

All submitted manuscripts must be in the journal's acceptable similarity index range:

≤ 20% – PASS; > 20% – REJECT.

International Standard Serial Number (ISSN)

An ISSN is an 8-digit code used to identify periodicals such as journals of all kinds and on all media–print and electronic.

Pertanika Journal of Science & Technology: e-ISSN 2231-8526 (Online).

Lag time

A decision on acceptance or rejection of a manuscript is reached in 90 days (average). The elapsed time from submission to publication for the articles averages 180 days.

Authorship

Authors are not permitted to add or remove any names from the authorship provided at the time of initial submission without the consent of the journal's Chief Executive Editor.

Manuscript preparation

For manuscript preparation, authors may refer to Pertanika's **INSTRUCTION TO AUTHORS**, available on the official website of *Pertanika*.

Editorial process

Authors who complete any submission are notified with an acknowledgement containing a manuscript ID on receipt of a manuscript, and upon the editorial decision regarding publication.

Pertanika follows a **double-blind peer-review** process. Manuscripts deemed suitable for publication are sent to reviewers. Authors are encouraged to suggest names of at least 3 potential reviewers at the time of submission of their manuscripts to *Pertanika*, but the editors will make the final selection and are not, however, bound by these suggestions.

Notification of the editorial decision is usually provided within 90 days from the receipt of manuscript. Publication of solicited manuscripts is not guaranteed. In most cases, manuscripts are accepted conditionally, pending an author's revision of the material.

The journal's peer review

In the peer-review process, 2 to 3 referees independently evaluate the scientific quality of the submitted manuscripts. At least 2 referee reports are required to help make a decision.

Peer reviewers are experts chosen by journal editors to provide written assessment of the **strengths** and **weaknesses** of written research, with the aim of improving the reporting of research and identifying the most appropriate and highest quality material for the journal.

Operating and review process

What happens to a manuscript once it is submitted to *Pertanika*? Typically, there are 7 steps to the editorial review process:

1. The journal's Chief Executive Editor and the Editor-in-Chief examine the paper to determine whether it is relevance to journal needs in terms of novelty, impact, design, procedure, language as well as presentation and allow it to proceed to the reviewing process. If not appropriate, the manuscript is rejected outright and the author is informed.

2. The Chief Executive Editor sends the article-identifying information having been removed, to 2 to 3 reviewers. They are specialists in the subject matter of the article. The Chief Executive Editor requests that they complete the review within 3 weeks.

   Comments to authors are about the appropriateness and adequacy of the theoretical or conceptual framework, literature review, method, results and discussion, and conclusions. Reviewers often include suggestions for strengthening of the manuscript. Comments to the editor are in the nature of the significance of the work and its potential contribution to the research field.

Pertanika Journal of Science & Technology
Pertanika Journal of Science & Technology

Pertanika Journal of Science & Technology
Pertanika Journal of Science & Technology

3. The Editor-in-Chief examines the review reports and decides whether to accept or reject the manuscript, invite the authors to revise and resubmit the manuscript, or seek additional review reports. In rare instances, the manuscript is accepted with almost no revision. Almost without exception, reviewers' comments (to the authors) are forwarded to the authors. If a revision is indicated, the editor provides guidelines for attending to the reviewers' suggestions and perhaps additional advice about revising the manuscript.

4. The authors decide whether and how to address the reviewers' comments and criticisms and the editor's concerns. The authors return a revised version of the paper to the Chief Executive Editor along with specific information describing how they have addressed' the concerns of the reviewers and the editor, usually in a tabular form. The authors may also submit a rebuttal if there is a need especially when the authors disagree with certain comments provided by reviewers.

5. The Chief Executive Editor sends the revised manuscript out for re-review. Typically, at least 1 of the original reviewers will be asked to examine the article.

6. When the reviewers have completed their work, the Editor-in-Chief examines their comments and decides whether the manuscript is ready to be published, needs another round of revisions, or should be rejected. If the decision is to accept, the Chief Executive Editor is notified.

7. The Chief Executive Editor reserves the final right to accept or reject any material for publication, if the processing of a particular manuscript is deemed not to be in compliance with the S.O.P. of *Pertanika*. An acceptance letter is sent to all the authors.

   The editorial office ensures that the manuscript adheres to the correct style (in-text citations, the reference list, and tables are typical areas of concern, clarity, and grammar). The authors are asked to respond to any minor queries by the editorial office. Following these corrections, page proofs are mailed to the corresponding authors for their final approval. At this point, **only essential changes are accepted**. Finally, the manuscript appears in the pages of the journal and is posted on-line.

Pertanika Journal of

SCIENCE & TECHNOLOGY

Vol. 33 (6) Oct. 2025

PERTANIKA JOURNALS

A scientific journal published by Universiti Putra Malaysia Press

PJST
Pertanika Journal of Science & Technology
AN INTERNATIONAL PEER-REVIEWED JOURNAL

UNIVERSITY PUBLICATIONS COMMITTEE

CHAIRMAN
Zamberi Sekawi

EDITOR-IN-CHIEF
Luqman Chuah Abdullah
Chemical Engineering

EDITORIAL STAFF
Journal Officers:
Navaneetha Krishna Chandran
Tee Syin Ying
Yavinaash Naidu Saravanakumar

Editorial Assistants:
Zulinaardawati Kamarudin

PRODUCTION STAFF
Pre-press Officers:
Ku Ida Mastura Ku Baharom
Nur Farrah Dila Ismail

WEBMASTER
IT Officer:
Kiran Raj Kaneswaran

EDITORIAL OFFICE
JOURNAL DIVISION
Putra Science Park
1st Floor, IDEA Tower II
UPM-MTDC Technology Centre
Universiti Putra Malaysia
43400 Serdang, Selangor Malaysia.

General Enquiry
Tel. No: +603 9769 1622
E-mail:
executive_editor.pertanika@upm.edu.my
URL: http://www.pertanika.upm.edu.my

PUBLISHER
UPM Press
Universiti Putra Malaysia
43400 UPM, Serdang, Selangor, Malaysia.
Tel: +603 9769 8855
E-mail: dir.penerbit@upm.edu.my
URL: http://penerbit.upm.edu.my

ASSOCIATE EDITOR
2024-2026

Miss Laiha Mat Kiah
Security Services Sn: Digital Forensic, Steganography, Network Security, Information Security, Communication Protocols, Security Protocols
Universiti Malaya, Malaysia

Saidur Rahman
Renewable Energy, Nanofluids, Energy Efficiency, Heat Transfer, Energy Policy
Sunway University, Malaysia

EDITORIAL BOARD
2024-2026

Abdul Latif Ahmad
Chemical Engineering
Universiti Sains Malaysia, Malaysia

Hsiu-Po Kuo
Chemical Engineering
National Taiwan University, Taiwan

Mohd. Ali Hassan
Bioprocess Engineering, Environmental Biotehnology
Universiti Putra Malaysia, Malaysia

Ahmad Zaharin Aris
Hydrochemistry, Environmental Chemistry, Environmental Forensics, Heavy Metals
Universiti Putra Malaysia, Malaysia

Ivan D. Rukhlenko
Nonliner Optics, Silicon Photonics, Plasmonics and Nanotechnology
The University of Sydney, Australia

Nor Azah Yusof
Biosensors, Chemical Sensor, Functional Material
Universiti Putra Malaysia, Malaysia

Azlina Harun@ Kamaruddin
Enzyme Technology, Fermentation Technology
Universiti Sains Malaysia, Malaysia

Lee Keat Teong
Energy Environment, Reaction Engineering, Waste Utilization, Renewable Energy
Universiti Sains Malaysia, Malaysia

Norbahiah Misran
Communication Engineering
Universiti Kebangsaan Malaysia, Malaysia

Bassim H. Hameed
Chemical Engineering: Reaction Engineering, Environmental Catalysis & Adsorption
Qatar University, Qatar

Mohamed Othman
Communication Technology and Network, Scientific Computing
Universiti Putra Malaysia, Malaysia

Roslan Abd-Shukor
Physics & Materials Physics, Superconducting Materials
Universiti Kebangsaan Malaysia, Malaysia

Biswajeet Pradhan
Digital image processing, Geographical Information System (GIS), Remote Sensing
University of Technology Sydney, Australia

Mohd Shukry Abdul Majid
Polymer Composites, Composite Pipes, Natural Fibre Composites, Biodegradable Composites, Bio-Composites
Universiti Malaysia Perlis, Malaysia

Sodeifian Gholamhossein
Supercritical technology, Optimization, nanoparticles, Polymer nanocomposites
University of Kashan, Iran

Ho Yuh-Shan
Water research, Chemical Engineering and Environmental Studies
Asia University, Taiwan

Mohd Zulkifly Abdullah
Fluid Mechanics, Heat Transfer, Computational Fluid Dynamics (CFD)
Universiti Sains Malaysia, Malaysia

Wing Keong Ng
Aquaculture, Aquatic Animal Nutrition, Aqua Feed Technology
Universiti Sains Malaysia, Malaysia

INTERNATIONAL ADVISORY BOARD
2024-2027

Hiroshi Uyama
Polymer Chemistry, Organic Compounds, Coating, Chemical Engineering
Osaka University, Japan

Mohini Sain
Material Science, Biocomposites,Biomaterials
University of Toronto, Canada

Mohamed Pourkashanian
Mechanical Engineering, Energy, CFD and Combustion Processes
Sheffield University, United Kingdom

Yulong Ding
Particle Science & Thermal Engineering
University of Birmingham, United Kingdom

ABSTRACTING AND INDEXING OF PERTANIKA JOURNALS

Pertanika Journal of Science & Technology is indexed in Journal Citation Reports (JCR-Clarivate), SCOPUS (Elsevier), EBSCO, ASEAN Citation Index, Microsoft Academic, Google Scholar and MyCite.

The publisher of Pertanika will not be responsible for the statements made by the authors in any articles published in the journal. Under no circumstances will the publisher of this publication be liable for any loss or damage caused by your reliance on the advice, opinion or information obtained either explicitly or implied through the contents of this publication.
All rights of reproduction are reserved in respect of all papers, articles, illustrations, etc., published in Pertanika. Pertanika provides free access to the full text of research articles for anyone, web-wide. It does not charge either its authors or author-institution for refereeing/publishing outgoing articles or user-institution for accessing incoming articles.
No material published in Pertanika may be reproduced or stored on microfilm or in electronic, optical or magnetic form without the written authorization of the Publisher.
Copyright ©2021 Universiti Putra Malaysia Press. All Rights Reserved.

Pertanika Journal of Science & Technology
Vol. 33 (6) Oct. 2025

Contents

Foreword i
Luqman Chuah Abdullah

Considerations of Earth Resistance Reduction Agent Studies 2395
Muhammad Noh Hamzah, Mohd. Zainal Abidin Ab Kadir, Shafreeza Sobri, Chandima Gomes and Syahrun Nizam Md. Arshad Hashim

Review Article 2413
Unveiling Insights on Scientometric Exploration of Trends in Power Cable Research – A Review
Khairunnisa Norfaeza Mohamed Khairul Fariz, Nik Hakimi Nik Ali, Mohd Iqbal Mohd Noor, Dalina Johari, Azrul Mohd Ariffin and Mohd Azraei Pangah @ Pa'at

Review Article 2441
A Comprehensive Review of Tuberculosis Detection and Prediction Using Hybrid Machine Learning Models on Ziehl-Neelsen-Stained Microscopy Images
Nur Amirah Kamaluddin, Murizah Kassim, Nor Salmah Bakar and Shuria Saaidin

Extraction and Characterisation of Natural Fibres from *Pennisetum purpureum*: Influence of Harvesting Age on Chemical, Mechanical, and Thermal Properties 2465
Mohd Fairus Kayat, Ridhwan Jumaidin, Mastura Mohamad Taha and Fahmi Asyadi Md Yusof

Harnessing Co-Culture of Hydrolase–Producing *Bacillus* spp. to Boost Kitchen Waste Biodegradation 2491
Nurhidayah Ramlee, Muhamad Aidilfitri Mohamad Roslan, Helmi Wasoh and Nor Aini Abdul Rahman

The Effects of Pre-Composting Subjected to Black Soldier Fly Larvae Digestion on Agri-Food Waste Maturation Indicators 2521
Muhammad Heikal Ismail, Naweedullah Amin, Anie Yulistyorini, Ridwan Muhamad Rifai, Khursheed Alam Khan and Luqman Chuah Abdullah

State-of-the-Art CNN Models for Plant Disease Classification: A Comparative Study 2541
Tanko Daniel Salka, Marsyita Hanafi, Sharifah M. Syed Ahmad Abdul Rahman and Dzarifah Mohamed Zulperi

Review Article 2565
Exploring Membrane Pervaporation Separation: A Review on Obtaining Combustible Ethanol and Butanol from Fermentation Broths
Jorge Arturo Moreno Cascante, Jairo Antonio Cubillos Lobo and Felipe Bustamante Londoño

Optimisation and Predictive Modelling of Natural Frequencies on Carbon/Glass Hybrid Composite Laminates 2589
Abdul Azim Taredi, Siti Mariam Abdul Rahman, Mohd Nor Azmi Ab Patar and Jamaluddin Mahmud

Case Study 2615
Mid-to-Post-Pandemic Mobility Patterns: A Case Study of Malaysia Highlighting Differential Trends in Work and Non-Work Trips
Megat-Usamah Megat-Johari, Nusayba Megat-Johari, Nischal Gupta, Hisham Jashami and Teik Hua Law

Review Article 2641
Credit Card Fraud Detection Using Machine Learning and Deep Learning Techniques: A Review
Shuchita Sheokand and Sunita Beniwal

Impact of Different Drone Sprayer Nozzle Types on Bio-Fertilizer Application in Wetland Paddy Cultivation 2665
Emmanuel Otache Adanu, Ahmad Suhaizi Mat Su, Azmi Yahya, Siti Amni Ismail, Nik Norasma Che'Ya, Muhammad Huzaifah Mohd Roslim, Abdul Shukor Juraimi, Muhamad Noor Hazwan Abd Manaf and Mst. Motmainna

Review Article 2685
Time Series Prediction for Cycling Sport Performance Through a Machine Learning Predictor
Muhammad Zulazri Hanis Mohd Nawi, Sukhairi Sudin, Fathinul Syahir Ahmad Saad, Muhamad Khairul Ali Hassan, Faranadia Abdul Haris, Nurul Syahirah Khalid and Kamarulzaman Kamarudin

Comparative Study of Nonlinear Regression Models for Cleanliness Factor Prediction in Coal-Fired Utilities 2723
Nur Aina Mohammad Abdul Aziz, Azura Che Soh, Asnor Juraiza Ishak, Nor Mohd Haziq Norsahperi and Amir Redzuan Mohd Ibrahim

Development and Verification of an MCNP5-Based Gamma Camera for ^{177}Lu SPECT Imaging 2747
Puvanesuawary Morthy, Marianie Musarudin, Nor Shazleen Ab Shukor, Mohamad Aminudin Said, Dong Xianling and M. Iqbal Saripan

Performance Evaluation of Jatropha Polyurethane Membrane Modified with Bio-Based Carbon Dots for Pb(II) Removal 2771
Umar Adam Majid, Zurina Zainal Abidin, Nur Haninah Harun and Khairul Faezah Md Yunos

RGB vs. HSV for Kitchen Fire Detection with YOLOv5 2797
Norisza Dalila Ismail, Rizauddin Ramli and Mohd Nizam Ab Rahman

Machine Learning Enabled IoT System for Agricultural Land Recommendation 2833
Bayu Rima Aditya, Anindia Agusta Ken Nadila, Muhammad Qanit Al-Hijran, Muhammad Bintang Ramadhan and Yudha Ginanjar

Enhancing Extractive Text Summarization Using Semantic-Based Similarity Matching 2855
Junjie Yang, Keng Hoon Gan and Jun Wang

Review Article 2881
Interaction Strategies in Augmented Reality-Situated Visualization for Decision Support: A Systematic Literature Review
Ainnecia Yoag, Masitah Ghazali and Farhan Mohamed

Foreword

Welcome to the sixth issue of 2025 for the Pertanika Journal of Science and Technology (PJST)!

PJST is an open-access journal for studies in Science and Technology published by Universiti Putra Malaysia Press. It is independently owned and managed by the university for the benefit of the world-wide science community.

This issue contains 20 articles: one case study article; six review articles; and the rest are regular articles. The authors of these articles come from different countries namely Afghanistan, China, Colombia, India, Indonesia, Malaysia, Nigeria, South Africa and United States.

A selected article titled "Extraction and Characterization of Natural Fibers from *Pennisetum purpureum*: Influence of Harvesting Age on Chemical, Mechanical, and Thermal Properties" examined new bio-based, natural cellulosic fibers from the *P. purpureum* plant, determining the optimal harvesting age at 30, 45, and 60 days. These fibers were subsequently analyzed to evaluate their physical, mechanical, thermal, and chemical properties. The relatively low density suggests that *P. purpureum* fibers could be advantageous in producing lightweight composite materials. The fiber of *P. purpureum* showed significant changes in chemical composition with increasing harvesting age. The mechanical properties of *P. purpureum* fibers were not solely dependent on cellulose content but were greatly influenced by the balance and interaction between cellulose and lignin. The results indicated that *P. purpureum* fiber has strong potential as a reinforcing material for thermoplastic green composites. Further detailed information can be found on page 2465.

The regular article entitled "Harnessing Co-Culture of Hydrolase-Producing *Bacillus* spp. to Boost Kitchen Waste Biodegradation" investigates the use of hydrolase-producing bacteria (HPB) co-cultures to enhance the biodegradation of kitchen waste (KW). The screening of 26 bacterial strains from fermented vegetables, agricultural soils, and in-house glycerol stock collections revealed *Bacillus licheniformis* 2D55 and *Bacillus xiamenensis* Y7 as standout candidates, exhibiting multiple hydrolase activities, including amylase, protease, lipase, and cellulase. In co-culture fermentation, these strains demonstrated superior hydrolase activities, microbial count, percentage of biodegraded total solids, and gross degradation rate during solid-state fermentation compared to monocultures and controls. These findings underscore the potential of HPB co-cultures as an effective

strategy to improve KW biodegradation, contributing to sustainable organic waste management solutions. Further details of this study can be found on page 2491.

Jorge Arturo Moreno Cascante and his teammates from the Pedagogical and Technological University of Colombia focused on the performance of pervaporation membranes, primarily polydimethylsiloxane-based membranes, for the separation of alcohols from fermentation broths. Additionally, the effects of composition and temperature are evaluated. Factors influencing process efficiency are also analyzed, and emerging technologies, new bacterial strains, and biomass sources are highlighted as strategies to enhance alcohol fuel production. The results showed that pervaporation has significantly improved biofuel production. Full details of this study are available on page 2565.

We anticipate that you will find the evidence presented in this issue to be intriguing, thought-provoking and useful in reaching new milestones in your own research. Please recommend the journal to your colleagues and students to make this endeavour meaningful.

All the papers published in this edition underwent Pertanika's stringent peer-review process involving a minimum of two reviewers comprising internal as well as external referees. This was to ensure that the quality of the papers justified the high ranking of the journal, which is renowned as a heavily-cited journal not only by authors and researchers in Malaysia but by those in other countries around the world as well.

We would also like to express our gratitude to all the contributors, namely the authors, reviewers and Editorial Board Members of PJST, who have made this issue possible.

PJST is currently accepting manuscripts for upcoming issues based on original qualitative or quantitative research that opens new areas of inquiry and investigation.

Editor-in-Chief
Luqman Chuah Abdullah

Pertanika J. Sci. & Technol. 33 (6): 2395 - 2412 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Considerations of Earth Resistance Reduction Agent Studies

Muhammad Noh Hamzah[1]*, Mohd. Zainal Abidin Ab Kadir[1], Shafreeza Sobri[1], Chandima Gomes[2] and Syahrun Nizam Md. Arshad Hashim[3]

[1]Department of Electric and Electronics Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[2]School of Electrical and Information Engineering, University of the Witwatersrand, 1 Jan Smuts Avenue, Braamfontein 2000, Johannesburg, South Africa
[3]Centre of Excellence for Renewable Energy, Faculty of Electrical Engineering Technology, Pauh Putra Main Campus, Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia

ABSTRACT

Earth resistance reduction agent (ERRA) studies are done to determine the best-performing material in different orientations and arrangements of earthing systems. Most studies in the past have adopted the installation of a row of earthing systems with ERRAs, using a single reference earthing. This is plausible only if the soil is highly homogeneous in its horizontal breadth and vertical depth soil resistivity attributes, so the comparisons can be made justly and accurately. This paper will demonstrate, using earth soil resistivity and electrode resistance values, that assuming a homogeneous soil in both horizontal and vertical directions can lead to misleading interpretations of the ERRA studies. A combination of earthing electrodes that are untreated and treated with ERRAs is used to demonstrate this issue. Soil resistivity and earthing resistance measurements were done for comparison purposes. Results found that the soil in the tested premises is inherently inhomogeneous. Earth resistance values of untreated soil were found to be 127.90, 45, 46.26, 35.84, 115.11, 51.95, 60.95, and 41.16 Ω, respectively. In comparison, earth resistance values of treated soil were found to be 617.33, 178.91, 163.75, and 148.23 Ω, respectively. Suggestions to effectively improve ERRA studies are then outlined in this paper, too.

ARTICLE INFO

Article history:
Received: 16 September 2024
Accepted: 17 March 2025
Published: 25 September 2025

DOI: https://doi.org/10.47836/pjst.33.6.01

E-mail addresses:
muhammadnoh.work@gmail.com (Muhammad Noh Hamzah)
mzk@upm.edu.my (Mohd. Zainal Abidin Ab Kadri)
shafreeza@upm.edu.my (Shafreeza Sobri)
chandima.gomes@wits.ac.za (Chandima Gomes)
syahrun@unimap.edu.my (Syahrun Nizam Md. Arshad Hashim)
* Corresponding author

Keywords: Earth resistance, inhomogeneous soil, soil resistivity

INTRODUCTION

Lightning and power systems need efficient earthing systems to return electrical current from the source or earth (Mohamad Nasir et al., 2021). Apart from the electrical integrity

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

(due to damage from corrosion), one of the other requirements of efficient electrical earthing is the earthing electrode dealing with its resistance when encased in its associated soil. The resistance to earth of an earthing electrode responsible for this purpose must be at its lowest according to the requirements of the standards (i.e., 1-10 Ω). Earthing resistance is proportional to soil resistivity and vice versa according to the Institute of Electrical and Electronics Engineers (IEEE) 81-2012, which indicates soil resistivity should be low to lower earthing resistances of earthing systems (Dilushani et al., 2020; IEEE, 2012). Earthing resistance with soil resistivity is an important parameter for engineers in ensuring electrical safety and earthing design (Al-Shawesh et al., 2021; Sinchi-Sinchi et al., 2022). Earth resistance determines the effectiveness of the earthing system that was designed efficiently to dissipate fault currents or lightning to the earth safely. This will reduce the risk of electric shocks, equipment malfunctions, and fire as the voltage build-up in the electrical system is minimal. However, the surroundings of soil and environment associated with a nearby structure that is to be protected by the earthing systems are already predetermined by its soil resistivity. The surrounding soil of earthing systems deployed can be altered in terms of its soil material by soil conditioning, enhancing, or treatment. A practical approach to reducing the ground resistivity of an earthing rod is to use ERRAs (Hamzah et al., 2023; Sinchi-Sinchi et al., 2022). IEEE 81-2012, IEEE 80-2013, and International Electrotechnical Commission (IEC) 62561-7 permit the use of ERRAs for this purpose (IEC, 2011; IEEE, 2012, 2015). In the technical field, ERRAs play a key role in improving the efficiency of earthing systems by lowering the resistance between the earthing and the earth. The ERRA efficiency parameter is given by:

$$\eta = 1 - \frac{R_e}{R_r} \quad [1]$$

where η represents the efficiency of an earthing electrode, R_e in reference to the earthing electrode, R_r. Higher η values portray higher efficiencies of the applied ERRA to the earthing system and vice versa, which in turn ensures dissipation of current in the case of fault, impulse, and stray currents that are significant in maintaining the safety and reliability of any electrical system (Androvitsaneas et al, 2017). Additionally, the parameter is crucial for engineers to evaluate the suitability of ERRAs in site-specific areas. In the economic sense, higher η values contribute towards cost savings in material and installation, maintenance reduction, space optimization, and infrastructure protection.

ERRAs can accommodate multiple earthing rods, earthing rod depth, and high soil resistivity in space-restricted areas. With increasing urbanization and population density (Patra et al., 2018), land becomes increasingly scarce; thus, a more sustainable approach to electrical earthing is required (Toh & Leong, 2023). This approach alters part of the

earth of the surrounding earthing rod by replacing it with ERRAs to enhance its soil resistivity in favor of electrical earthing (IEEE, 2015). This means that the soil resistivity of the enhanced area will be lowered without too much land space being used as opposed to without ERRAs. Consequently, the earthing resistance of the associated earthing rod of a certain area will be lowered as well to a certain amount, depending on the efficiency of the ERRA used and the surrounding soil profile. As a result, a lower number of earthing electrodes is used to reach the low value of earthing resistance (IEEE, 2012). Areas where high resistivity of soil is prevalent are suitable for this kind of solution. One advantage of this method is that when dealing with soils that are hard for the electrodes to penetrate, such as in the study of Sinchi-Sinchi et al. (2022), the soil of the area can be altered with ERRAs that are generally easy for earthing electrodes to be driven in or, in some cases, installed horizontally depending on the soil profile.

Swileam et al. (2019) showed that the soil resistivity is unique to its own place. Various factors affect the soil resistivity at different depths and breadths, making it diverse at every point (Gerscovich & Vipulanandan, 2023). It is essential to recognize that failing to examine the soil profile carefully can lead to inaccuracies in ERRA studies or the design of earthing systems. A lot of assumptions and ambiguity are made this way, which gives a skewed interpretation of the soil profile. The study showed that there is a significant difference between the average and localized soil resistivity at different places, even at distances of 5 meters apart. Waste of potential soil space can also occur when the soil resistivity of an area can be utilized for an earthing system designed to go unnoticed in the area due to poor electrical resistivity surveying (Sinchi-Sinchi et al., 2022). Therefore, it is essential that before earthing systems are driven into the ground, their respective localized soil resistivity is determined so earthing arrangements can be designed as efficiently as possible.

At places where high soil resistivity is measured, alterations of the soil can be done efficiently in terms of cost and management, especially when there is limited space. As the depth of the soil increases, soil resistivity decreases (Lech et al., 2020), although this might not necessarily be the case. This information is critical for ERRA studies, as the alteration of the soil may only require treatment of the topsoil, depending on the soil profile, which is less costly. There has not been any standard procedure or proper methodology to test ERRAs in a soil area. Typical soil resistivity profile surveys are shown in Figure 1, which is a scattered soil resistivity measurement survey compared to a localized one. Figure 1 a) shows soil resistivity measurement to obtain the soil resistivity in several areas of the field. Figure 1 b) shows localized soil resistivity in a particular area. Figure 1 c) shows a soil resistivity survey in several localized areas. Localized soil resistivity surveys are done to see the relevant soil resistivity values across the row of earthing electrodes that are to be tested with the ERRAs. This approach is more effective than a scattered measurement for observing soil homogeneity.

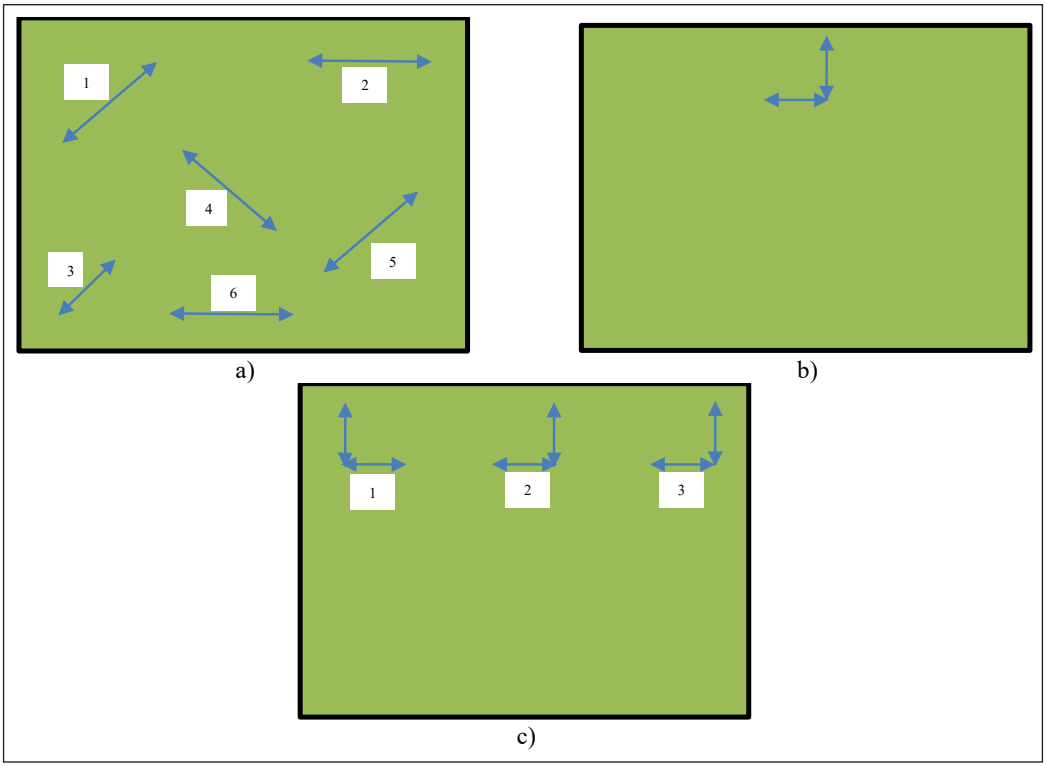


Figure 1. a) Average soil resistivity survey in a larger surface area at several spots of the field; b) Localized soil resistivity measurement in different directions of the field; and c) Localized soil resistivity in several areas of the same vertical plane
Note. The number of measurements in a particular space can vary

A soil resistivity survey is done to search for a high soil resistivity area for ERRA studies based on the horizontal and vertical profiling of the respective soil (Mostafa et al., 2018). Usually, the soil is assumed to be homogeneous across the whole field based on the survey, when this is highly unlikely due to having different soil structures across the soil bed (C. Chen et al., 2024; Zhang et al., 2019). A high soil resistivity field is suitable for this kind of study, as it would test the effectiveness of ERRAs in reducing the earth resistance of an earthing electrode driven into the earth. To see the resistance to earth of an electrode after the soil survey, a rod is usually driven at a place where it is suited to be the reference electrode, usually at a corner or end of a given field, such as in Figure 2. The rest of the earthing electrodes with ERRAs are deployed along the row initiated by the reference electrode. Most ERRA studies in the past have been evaluated based on one or two reference electrodes (Hamzah et al., 2023; S.-D. Chen et al., 2006; Wan Ahmad et al., 2018). An earthing electrode is usually placed at the start or end of a row of earthing with ERRAs, with the rest of the earthing electrodes enhanced with their respective ERRAs, as demonstrated in Figure 2. Most of the ERRA studies in the past have been done this way.

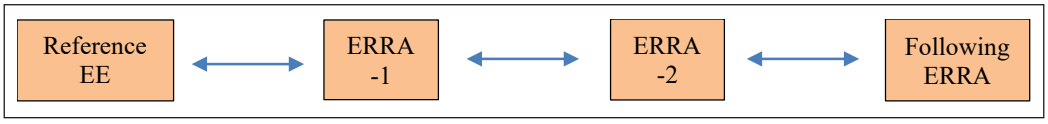


Figure 2. Typical earth resistance reduction agent (ERRA) study of row configuration placement with their respective earthing electrodes (EE)

Soil Inhomogeneity in ERRA Studies

The difference between the reduced earthing resistance of the earthing rod with ERRAs and the reference earthing is the efficiency of the ERRA in reducing earth resistance values. Fluctuation values over time are also a measure of the reliability of the ERRAs. Measurements are taken for an extended period at different frequencies to evaluate their performance in terms of fluctuation and earth resistance decrement. Androvitsaneas et al. (2012) studied the earth resistances of 5 earthing systems with ERRAs and introduced the efficiency factor to evaluate their performances for a period of one year. This is similar to the one used by L.-H. Chen et al. (2009) and S.-D. Chen et al. (2006), as mentioned in Trifunovic (2018). The ERRA efficiency formula is given in Equation 1.

The performances of their earthing systems were found not to be consistent throughout the one-year study but showed a decrease with respect to the reference earthing electrode. Though the efficiency factor proposed by Androvtsaneas et al. (2012) gives some insight into the performance of the earthing systems, the information that is obtained from the efficiency factor can be misleading and would only work for homogeneous soils in their horizontal lengths and vertical depths, which is highly improbable (Lech et al., 2020). This method is viable only if there is a certainty that for every electrode placed in the ground, it has the same or at least the same soil resistivity values, which is difficult to determine even with the aid of measuring the local soil resistivity, which will be demonstrated later. Hence, usually the soil resistivity is sought to determine the average soil resistivity around the area of interest. However, only determining the soil resistivity of a respective area may not suffice to interpret the homogeneity of the adopted soil if done in a scattered manner, as mentioned previously. The placement configuration done by S.-D. Chen et al. (2006) would be a more accurate kind of ERRA study since two earthing electrodes cover the area where the ERRA in their study are being tested. This testing configuration is shown in Figure 3. The earthing resistance values of the reference electrodes differed by a certain value in each periodic measurement, but not significantly.

Most of the studies have tested different ERRAs with their earthing electrode to be tested against the reference earthing. ERRAs can be of bentonite, marconite, and fly ash (Mohd Tadza et al., 2019). This means for every type of ERRA encased to one earthing electrode tested against the reference earthing, there will only be one earthing electrode with

one kind of ERRA that is tested. If the soil is inhomogeneous to begin with, an incorrect account of their efficiencies will be displayed. This will be demonstrated in the results section, where three earthing electrodes with the same kind of ERRA are tested against a reference electrode, which was found to be different in efficiency levels.

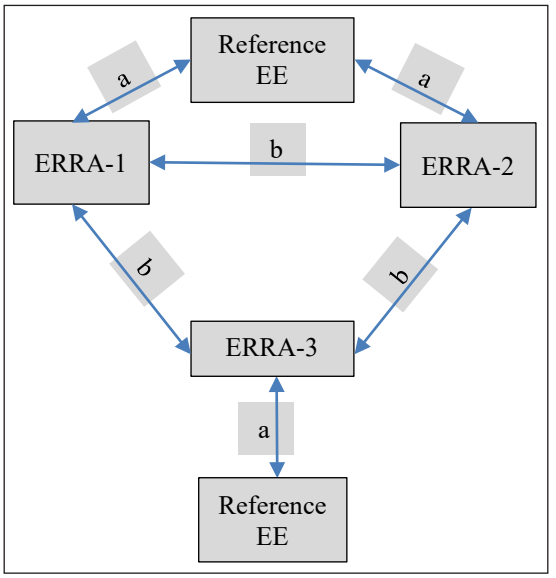


Figure 3. Two reference earthing covering the ERRA electrodes in between (Adopted from S.-D. Chen et al., 2006)
Note. EE = Earthing electrode; ERRA = Earth resistance reduction agent; a = Distance between tested reference earthing and tested earthing electrode; b = Distance between tested earthing electrodes

METHODOLOGY

The ERRA studies typically begin by determining the average soil resistivity of an area. High soil resistivity areas are preferable to see the extent of the effectiveness of ERRAs in decreasing earth resistance values. To demonstrate that ERRA studies can be inaccurate without taking proactive measures in inhomogeneous soils, copper-bonded earthing electrodes of dimensions shown in Figure 4 were installed at eight different locations in the same field, as seen in Figures 5 and 6. In Figure 4, the reference earthing is simply the tested earthing conductor with the native soil, while the industrial earthing system is encased with the industrial ERRA.

The distances of 7 m and 9 m in Figure 6 were chosen based on the site-specific limitations due to restricted space in the area, so the earthing electrodes can be spread as far as possible to evaluate the variety of the earthing resistance soil profile of the area using the eight earthing electrodes. The standards also suggest that the distances between electrodes should be at least equal to their length to avoid electrical interference, where the distances are at least 4.7 times their earthing electrode length (IEEE, 2000). The earthing electrodes

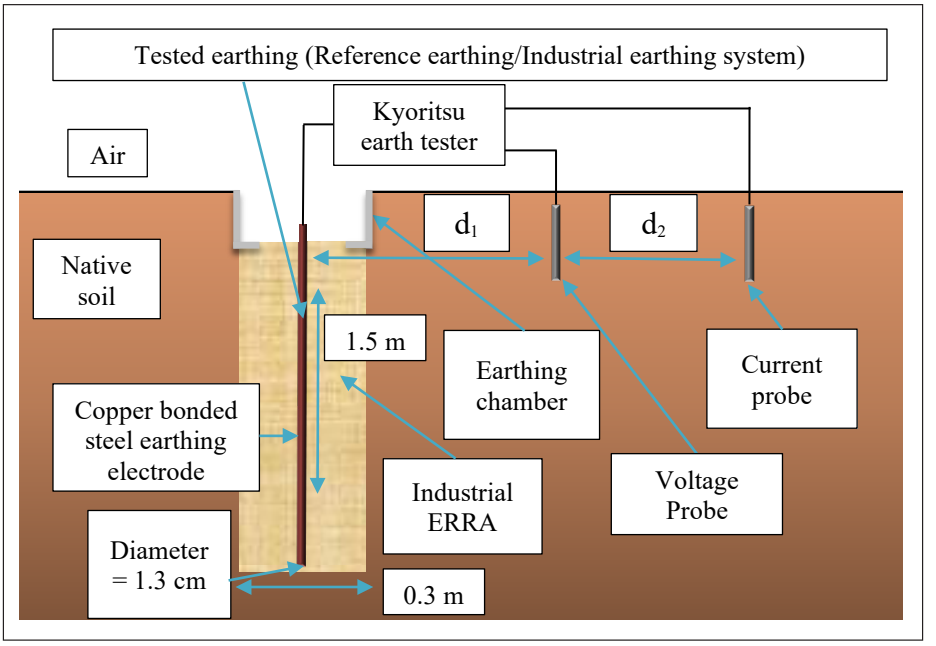


Figure 4. Dimensions of earthing electrodes in the experiment and measuring earthing electrode resistance
Note.
d_1 = The distance from the tested earthing electrode to the voltage probe
d_2 = The distance from the voltage probe to the current probe
ERRA = Earth resistance reduction agent



Figure 5. One of the untreated earthing electrodes (EE-3)

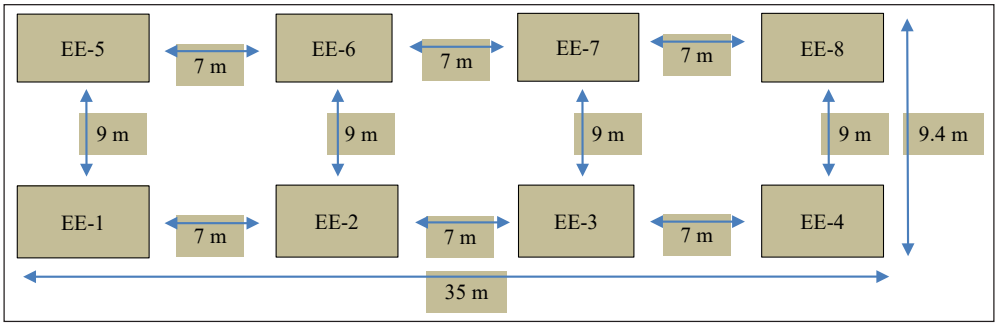


Figure 6. Untreated earthing electrode (EE) placement in the same field (aerial view)

in the field are untreated with any ERRA, which means the earthing resistances are of the native soil. Average soil resistivity values were also measured at a depth of 2 m for a period of a year using the Megger soil resistivity tester (United Kingdom). Earth resistances were measured using the fall-of-potential method for a duration of 19 weeks every other week (second week of every two weeks) using the Kyoritsu earthing measurement tester (Japan) as shown in Figure 4 using the fall-of-potential method. All measurements were taken according to the IEEE 80-2013 and IEEE 81-2012 standards, where Ohm's Law (V = IR) is used to calculate earthing resistance, which is computed in the Kyoritsu earth resistivity tester. The efficiency of ERRAs is computed using Equation 1.

Another earthing system was deployed in a different location within the premises of Universiti Putra Malaysia (UPM), and it will also be considered in this study. They include the industrial earthing systems that were deployed in a high soil resistivity area, which can be seen in Figures 7 and 8. The ERRA that is used in this study is an industrial-grade material named Earth-Enhancing Material (EEM)-10. This separate study tests a reference earth and three earthing electrodes, fully encased with EEM-10. The dimension of the configuration is given in Figure 4. Soil resistivity measurements were also taken in the respective local earthing of 1 meter and 2 meters depth, as given in Table 1, to see the inhomogeneity of the soil.

Table 1
Performance of industrial earth resistance reduction agent (ERRA) in inhomogeneous soil of treated earth

Earthing electrode	Average (Ω)	Fluctuations (Ω)	Average drop (%)	Respective soil resistivity	
				1 meter (Ω-m)	2 meters (Ω-m)
Reference electrode	617.33	72.65	-	1650.91	907.92
Industrial ERRA-1	178.91	21.93	71.01	1883.38	818.38
Industrial ERRA-2	163.75	19.85	74.47	1690.18	783.83
Industrial ERRA-3	148.23	16.85	75.99	1529.96	753.98

Note. Fluctuation is categorized by standard deviation



Figure 7. An industrial earth resistance reduction agent earthing (treated earthing)

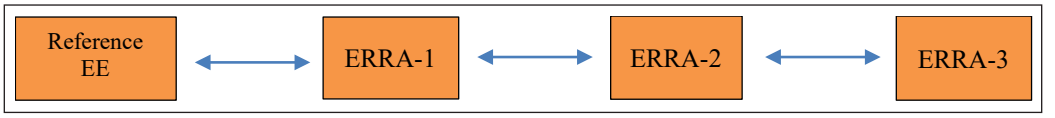


Figure 8. Industrial earth resistance reduction agent (ERRA) configuration (treated earthing electrodes)

RESULTS AND DISCUSSION

Earth Resistance in Inhomogeneous Soil with Untreated Earth

As can be seen in Table 2 in relation to Figure 6, the earthing resistance values are not consistent, given that they are in the same field and untreated with ERRAs. An ideal earth resistance value for ERRA studies should be around the values of EE-1 or EE-5 for reference earthing placed in the field due to their high values compared to the tested electrodes (EE-2, EE-3, EE-4, EE-6, EE-7, and EE-8). The average earthing resistances for Earthing 1 to 8 were found to be 127.90, 45, 46.26, 35.84, 115.11, 51.95, 60.95, and 41.16 Ω, respectively, as can be seen in Figure 9. The results almost imitate an ERRA study, where EE-2, EE-3, EE-4, EE-6, EE-7, and EE-8 are seen to be lowered with respect to EE-1 and EE-5 of their respective rows when none of the EEs are treated with ERRAs. The highest and lowest average earth resistances are found from EE-1 and EE-4, having values of 127.90 and 35.84 Ω, respectively. It is clear from the measurements made that there is a variety of earthing resistance values that indicate soil inhomogeneity. This is as stated in IEEE STD81-2012, where driving earthing electrodes is indicative of the soil resistivity of a particular area along a lateral plane of the soil. Due to the soil texture and structures, various soil resistivity values exist (Gerscovich & Vipulanandan, 2023).

Table 2
Earth resistance and localized equivalent soil resistivity of each earthing electrode

Earthing electrode number	Average (Ω)	Standard deviation (Ω)	Localized equivalent soil resistivity (Ω-m)
1	127.89	12.77	206.85
2	45	4.61	72.78
3	46.26	0.99	74.82
4	35.84	0.9	57.97
5	115.11	6.26	186.18
6	51.95	4.13	84.03
7	60.95	2.48	98.58
8	41.16	1.64	66.57

Using soil resistivity (Equation 2), the localized equivalent soil resistivity for the respective earthing electrode can be deduced as in Table 2 (IEEE 81-2012) to see the soil resistivity variation.

$$\rho_{eq} = \frac{2\pi RL}{\left[\ln\left(\frac{8L}{d}\right) - 1\right]} \quad [2]$$

where ρ_{eq} represents the equivalent soil resistivity, R is the measured earthing resistance, L is the length of the electrode, and d is the electrode's diameter.

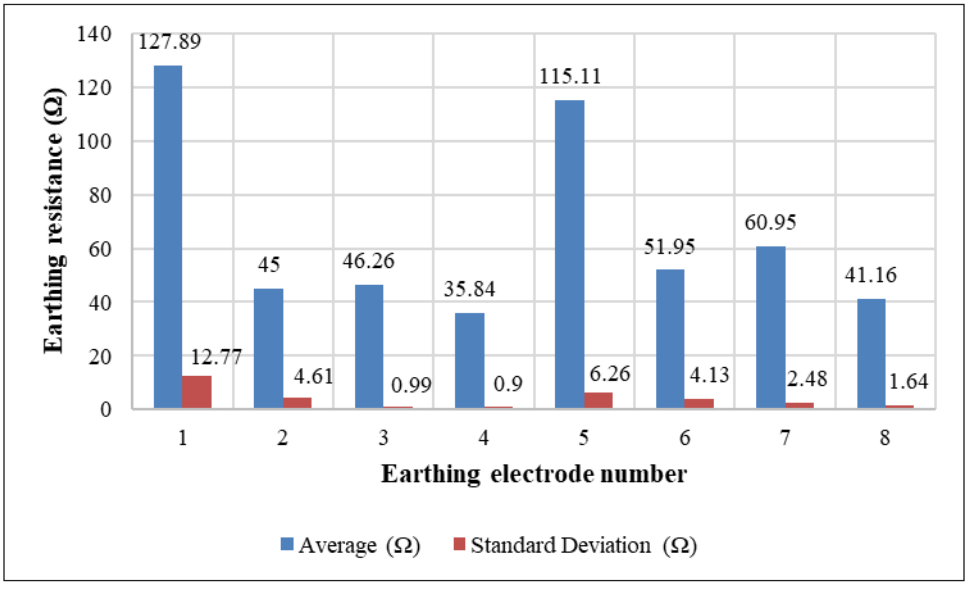


Figure 9. Variation of earthing resistances of untreated earthing electrodes on the same plane

As can be seen in Table 2, the localized soil resistivity varies depending on the area, as depicted in Figure 6, and its respective earth resistance electrodes. Figure 9 shows a comparison of the average earthing resistance, accompanied by its standard deviation, of the earthing electrodes in the field. Less deviation is seen in lower value earthing resistances from Table 2, which indicates that lower soil resistivity values are more likely to fluctuate less in earth resistance value with respect to time. This is highly important because earthing systems should be more stable in terms of fluctuation values. High fluctuation of earthing resistance values over time may indicate underlying problems such as soil moisture variations, corrosion or degradation of the electrodes, or loose connections.

When finding a reference electrode for an ERRA study, if the assumption is made based on the average soil resistivity and EE-1 as the reference, then EE-2, EE-3, and EE-4 would also be assumed to have about the same average earth resistance values as EE-1. The measured earth resistance values of EE-2, EE-3, and EE-4 have proved that this is not the case. The same would be for the case of EE-5 and the rest of its row. In this scenario, if ERRAs were installed in EE-2, EE-3, or EE-4 while having EE-1 as the reference, the percentage drop with the effect of ERRAs would be determined with the aid of the soil surroundings of the respective earthing electrodes. This will pose an inaccurate account of ERRA accuracy regarding the earth resistance drop. Also, if ERRAs were installed in areas where the earthing resistance was already lowered to begin with, it would be a waste of ERRA in practical terms. This is because if a reduction of earthing resistance were to happen to these local earthing electrode areas, there would only be a minor to no reduction. Hence, applying ERRAs in these areas would be an inefficient earthing system design in terms of estimating the efficiency of an ERRA.

The other scenario would be that when EE-4 or EE-8 is sought first as the reference, the earthing electrode value is already lowered due to the soil environment encasing their respective earthing electrode. This poses the problem of needing to remove the earthing electrode, as it is not suitable to be a reference electrode. However, it does provide an indication of whether an ERRA study is viable in the area. This also provides vital information that soil resistivity values are not enough to determine if an area is suitable for an ERRA study. This proves qualitatively that soils in this area of study are inherently inhomogeneous and not suitable for ERRA studies. However, in the case of real earthing applications, only earthing electrodes numbers 1 and 5 need to be treated with ERRAs. Hence, it is proven that inhomogeneous soils are not suitable for ERRA studies, where incorrect assumptions on the soil profile will lead to inaccuracy of the study.

Using information on the soil profile of the area, if the area is needed to be used to install an earthing system, only EE-1 and EE-5 may need ERRAs to reduce their earth resistance. The soil profile indicates that there is a soil inhomogeneity case that the designer of the earthing system can utilize. There is a decrease in soil resistivity value in horizontal and vertical directions, as indicated in Figure 9 (horizontal decrease in the

downward direction) and Table 2 (vertical decrease in the right direction). The earthing system can then be installed by connecting all the earthing electrodes present in the area to achieve the earthing resistance desired by the standard of around 1-10 Ω (IEEE, 2015). The area where lower earth resistance values are found may indicate that the area is high in moisture, and vice versa.

Figure 10 shows the 3D surface plot of the earthing resistance across electrode positions. The figure gives a better representation of the variability of earthing resistance across the plane. Using the coefficient of variability (*CV*) formula:

$$CV = \frac{\sigma}{\mu} \times 100\% \quad [3]$$

where σ and μ are the standard deviation and mean values of the untreated earthing resistance. The *CV* was found to be 54.18%, indicating that there is high variability of

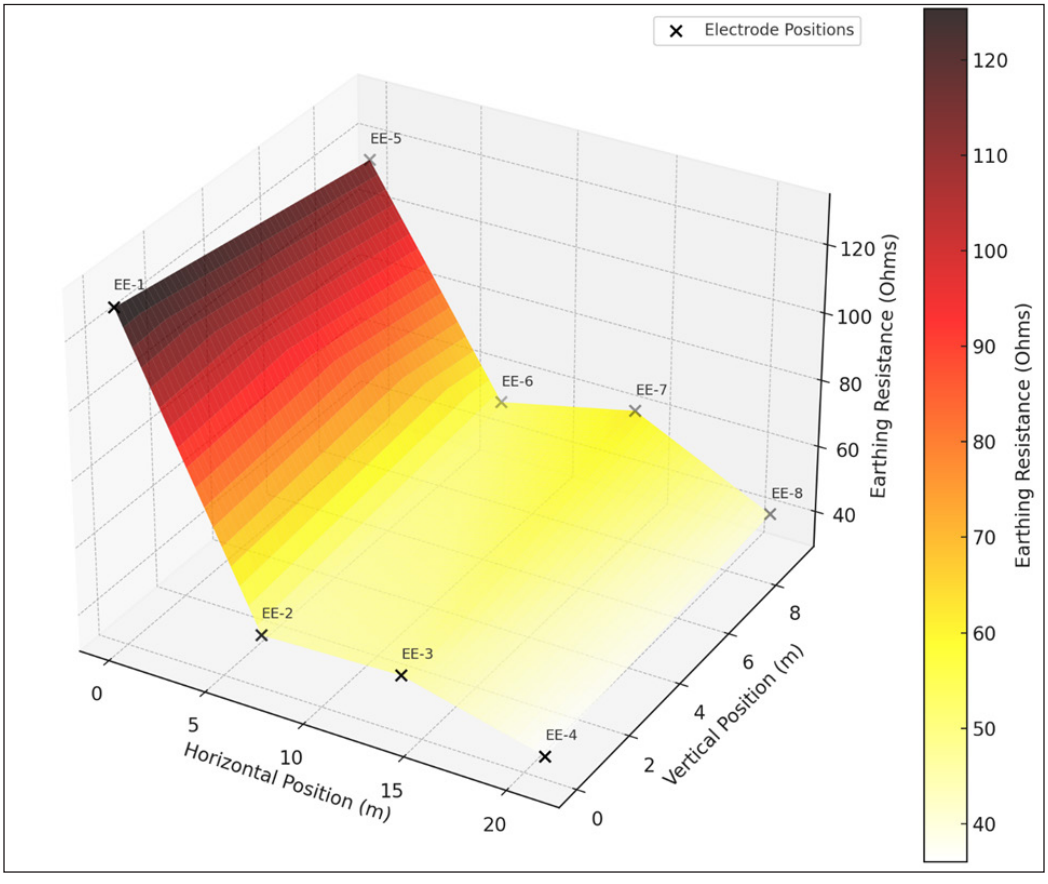


Figure 10. The 3D plot of untreated earthing electrodes (EE)

earthing resistance in the soil. From the figure, there is a very noticeable difference in soil variability, which reflects the inhomogeneity of the soil profile in the area. The darker regions (e.g., EE-1 and EE-5), where earthing resistance is higher, could indicate that there is lower moisture and salinity. However, brighter regions (e.g., EE-2, 3, 4 and EE-6, 7, 8) of the 3D plot show less variety with lower earthing resistance values, indicating higher levels of moisture and salinity. These variations of earthing resistance show that if ERRAs are to be tested in this region, the underlying resistivity variations may appear distorted than their actual value.

Earth Resistance in Inhomogeneous Soil of Treated Earth

In a separate study to see the earth resistance of industrial ERRA in a slightly inhomogeneous soil, the *CV* was found to be 9.38% using Equation 3. The variability calculation only included industrial earthing 1, 2, and 3, as they are the only ones treated with ERRAs. Table 1 and Figure 11 show the performance of the industrial ERRA done in a separate field. An ideal ERRA study would have a significant percentage drop regarding the reference electrode, as demonstrated in Figure 11. The ERRA efficiency is computed using Equation 1. The average earth resistances for the reference electrode, Industrial ERRA-1, Industrial ERRA-2, and Industrial ERRA-3 are 617.33, 178.91, 163.75, and 148.23 Ω, respectively. The average percentage drop of Industrial ERRA-1, Industrial ERRA-2, and Industrial ERRA-3 based on the reference drop is 71.01, 74.47, and 75.99%, respectively. Although the ERRAs installed at the respective electrodes are of the same type and similar amount, there is a slight percentage difference in the average drop.

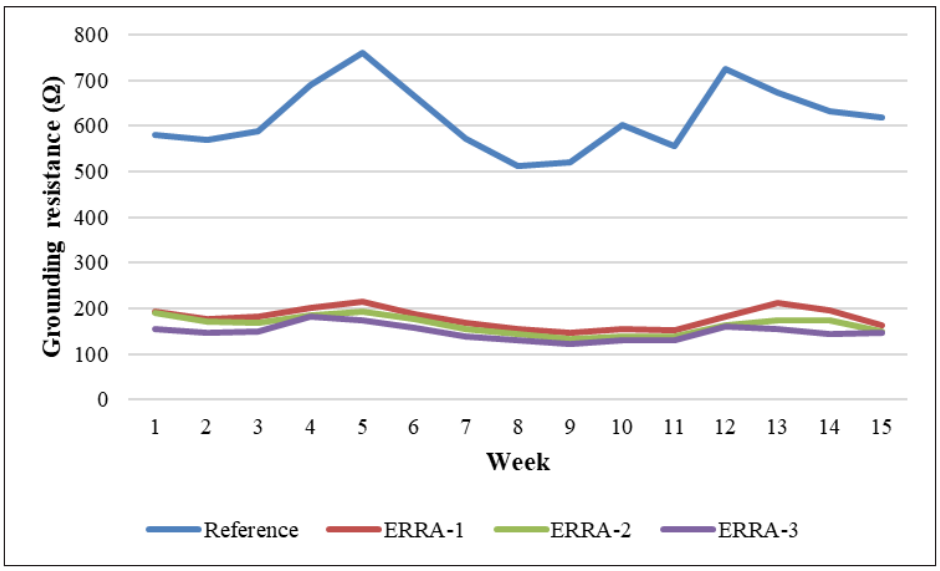


Figure 11. Performance of industrial earth resistance reduction agents (ERRAs)

This section of the study suggests that a perfectly homogeneous soil is impossible to achieve when the localized soil resistivity values slightly differ from one point to another. The reference, Industrial ERRA-1, Industrial ERRA-2, Industrial ERRA-3, and Industrial ERRA-4 soil resistivity measurements show the values of 1650.91, 1883.38, 1690.18, and 1529.96 Ω-m for 1 meter depth, and 907.92, 818.38, 783.83, and 753.98 Ω-m, respectively. This shows that the localized soil resistivity of each earthing electrode varies slightly from one another, which will result in different performances of the industrial ERRA with respect to their localized soil. This part of the study demonstrates that there is a variation in earthing resistance value, although the same ERRA was deployed to the three tested earthing electrodes. The surrounding soil resistivity plays a role in determining the final value of the earthing electrode displayed in Table 1.

Recommendations for Improved Accuracy of ERRA Studies

As shown in the results and discussion section, it is evident that ERRA studies may lead to inaccurate results if proper inspection is not done beforehand**.** It is evident from the results and literature that soils are inhomogeneous across the soil bed (Zhang et al., 2019). As a result, the earth resistance of earthing electrodes will vary depending on the surrounding soil that it is encased in. Thus, soil resistivity measurements should be taken accurately across the soil bed to reduce the inaccuracies of the ERRA studies. Depending on the soil profile, different configurations can be adopted for the ERRA study if precise and sufficient measurements are made. The row configuration, which is the conventional way, can be used if the soil resistivity survey shows a good sign. The experimental study or real-world applications of electrical earthing are site-specific and may require different approaches for optimal outcomes. Since it is almost impossible to attain a perfectly homogeneous soil, the ERRA studies will have to depend on the information obtained from the soil profile data.

It is recommended that the soil resistivity measurement be done specifically where the earthing electrode is going to be driven. This will give more confidence in deploying the earthing electrodes that are to be tested with their respective ERRA. While it may be more challenging, the expected outcome is going to be more precise and accurate. Regular monitoring is crucial to observe the fluctuations in values represented by the standard deviation value over time. An increase in standard deviation values over time may indicate that the integrity of the installed earthing electrode or system has been compromised. Thus, it is important that an effective earthing system has a low earthing resistance value according to the standard and can maintain this value for long periods of time.

Practical and Environmental Implications of Installing Earthing Systems

As there are practical implications of soil homogeneity in designing the electrical earthing system, the engineer must do an extensive soil resistivity survey and analysis to minimize

the time and effort required to install the earthing system. There are various considerations to be made that will provide the engineer with information to ease the process of designing and installing the earthing system. The engineer must be able to determine if the earthing system electrodes should be installed vertically, horizontally, or a combination of both, depending on the soil profile (IEC, 2002). Thus, engineers must also have environmental and geotechnical insights into the surrounding environment. This is highly important, especially when the cost of excavation of the soil is high. The soil profile does not consist only of the soil resistivity value, but also other signs of the nature of the surrounding soil. For instance, a high resistivity value measured at a depth of meters may indicate the presence of bedrock (Ward et al., 2014). Hence, it may not be viable to install the earthing system vertically in that area. Installing the earthing system horizontally may be a better option depending on the allowable space. However, in certain cases where the area is restricted, soil modifications (ERRAs) are needed to lower the soil resistivity.

Soil resistivity can also indicate if there is underground water at the bottom. Soils that are saturated with water have low soil resistivity as the soil is more conductive. Sudden changes to very low resistivity after certain depths indicate that groundwater is available (X. Chen et al., 2022). Thus, the electrical earthing engineering installation team may use this to their advantage. Longer earthing rods or conductors can be driven relatively more easily into the soil area, due to the low resistivity and softness of the soil compared to dry soils.

Soil resistivity can provide many indicators of the soil. Electrical engineers can use these markers to their advantage to save costs when designing earthing systems, as mentioned above. However, regarding the testing of ERRAs, the person in charge of the study must find a location that can accommodate these tests, as this will be vital to assess the performance of the ERRA. Hence, careful investigation of the soil profile must be done before accommodating for a slight difference in soil resistivity values of the tested earthing electrodes.

Another consideration that engineers may overlook is the electrical contact that the driven electrode has with the soil. There have been reports that electrodes encased with bentonite can shrink due to dry conditions (Mohd Tadza et al., 2019). This, in turn, results in decreased electrical contact with the associated electrode to the bentonite aggregate, which results in higher earth resistance. Soil erosion may occur in areas where the soil is sandy and high in water displacement. This can also result in reduced electrical contact of the earthing electrode with the associated soil, which increases the earthing resistance of the overall earthing system. Hence, it is appropriate that regular earth resistance measurement is made to maintain the integrity of the earthing system and make proper adjustments.

CONCLUSION AND RECOMMENDATIONS

The soil resistivity variability in inhomogeneous soil is impactful in determining the ERRA experiments. The nature of soil is different in terms of environmental conditions, soil composition, moisture content, and other factors. This can result in inaccurate assessments, which may lead to inefficient earthing systems being developed and increase safety risks.

This paper has demonstrated that an ERRA study can be misleading due to the wrong estimation of the soil resistivity profile around the soil bed that is being tested. Soil resistivity values can also give misleading interpretations of the earth resistance that an earthing rod can have after being driven into an earth area due to wrong assumptions. The information obtained in this study can be used as a guide to develop earthing systems with greater efficiency in terms of soil resistivity. This measure will save costs on electrode and land usage, which is another problem that arises.

It is proposed that ERRA studies deploy more than 1 reference earthing (of about 4-5) to each zone related to the tested earthing electrode with ERRA, if it is done using the row configuration at the allowable distance according to the standards. This is especially when there are a lot of ERRAs to be studied for their performances as depicted in Figure 3. More than 1 reference earthing allows for more accuracy of the baseline reference earthing value. This will allow for diverse soil conditions (a variety of earth resistance or soil resistivity values) to be standardized, ensuring that reference earthing results are generalized and consistent in baseline readings. Multiple reference earthing can capture the earth resistance variability of the soil profile of concern wherever it is deployed. Hence, a controlled environment consisting of multiple earthing reference electrodes should be considered for better accuracy in determining the efficiency of ERRAs.

ACKNOWLEDGEMENTS

We would like to extend our deepest gratitude to Prof. Ir. Dr. Ts. Mohd Zainal Abidin Ab Kadir and Dr. Shafreeza Sobri for their invaluable knowledge and guidance throughout this research. Their insights have significantly contributed to the successful completion of this paper. Our sincere thanks also go to Prof. Dr. Chandima Gomes for both his extensive knowledge and generous financial support. His contributions have been crucial to the advancement of this research. Lastly, we are grateful to Assoc. Prof. Dr. Syahrun Nizam Md Arshad Hashim for his meticulous review and constructive feedback on the manuscript, ensuring the rigor and clarity of our work.

REFERENCES

Al-Shawesh, Y., Lim, S. C., & Nujaim, M. (2021). Analysis of the design calculations for electrical earthing systems. *International Review of Electrical Engineering*, *16*(2), 104–117. https://doi.org/10.15866/iree.v16i2.16839

Androvitsaneas, V. P., Gonos, I. F., & Stathopulos, I. A. (2012). Performance of ground enhancing compounds during the year. In *International Conference on Lightning Protection* (pp. 1–5). IEEE. https://doi.org/10.1109/iclp.2012.6344356

Androvitsaneas, V. P., Gonos, I. F., & Stathopulos, I. A. (2017). Research and applications of ground enhancing compounds in grounding systems. *IET Generation, Transmission and Distribution*, *11*(13), 3195–3437. https://doi.org/10.1049/iet-gtd.2017.0233

Chen, C., Yang, Z., Deng, Y., Ma, H., & Qian, J. (2024). Experimental investigation of thawing behavior of saline soils using resistivity method. *Journal of Geophysics and Engineering*, *21*(3), 854-867. https://doi.org/10.1093/jge/gxae037

Chen, L.-H., Chen, J.-F., Liang, T.-J., & Wang, W.-I. (2009). A research on used quantity of ground resistance reduction agent for ground systems. *European Transactions on Electrical Power*, *20*(4), 408–421. https://doi.org/10.1002/etep.322

Chen, S.-D., Chen, L.-H., Cheng, C.-K., & Chen, J.-F. (2006). An experimental study on the electrical properties of fly ash in the grounding system. *International Journal of Emerging Electric Power Systems*, *7*(2). https://doi.org/10.2202/1553-779x.1284

Chen, X., Guo, Z., Liu, C., Liu, J., & Wu, Q. (2022). Groundwater detection using the pseudo-3D resistivity method: A history of case studies. *Applied Sciences*, *12*(13), 6788. https://doi.org/10.3390/app12136788

Dilushani, P. D. R., Nawodani, W. R. N., Bambaravanage, T., & Udayakumar, K. A. C. (2020). Soil resistivity analysis and earth electrode resistance determination. *IOSR Journal of Electrical and Electronics Engineering*, *15*(2), 26–35.

Gerscovich, D. M. S., & Vipulanandan, C. (2023). Data analyses to correlate the soil properties to the electrical resistivity. *Geotechnical and Geological Engineering*, *41*, 4507–4528. https://doi.org/10.1007/s10706-023-02529-y

Hamzah, M. N., Gomes, C., Afiq, M., & Ab Kadir, M. Z. A. (2023). Performance of active electrical earthing systems: Results and discussion. In *12th Asia-Pacific International Conference on Lightning* (pp. 12–15). IEEE. https://doi.org/10.1109/APL57308.2023.10182183

Institute of Electrical and Electronics Engineers. (2000). *IEEE guide for safety in AC substation grounding*. IEEE. https://doi.org/10.1109/IEEESTD.2000.91902

Institute of Electrical and Electronics Engineers. (2012). *IEEE guide for measuring earth resistivity, ground impedance, and earth surface potentials of a grounding system*. IEEE. https://doi.org/10.1109/IEEESTD.2012.6392181

Institute of Electrical and Electronics Engineers. (2015). *IEEE guide for safety in AC substation grounding*. IEEE. https://doi.org/10.1109/IEEESTD.2015.7109078

International Electrotechnical Commission. (2002). *IEC 60364-5-54: Electrical installations of buildings – Part 5-54: Selection and erection of electrical equipment – Earthing arrangements, protective conductors and protective bonding conductors*. IEC.

International Electrotechnical Commission. (2011). *IEC 62561-7: Lightning protection – Part 7: Requirements for earthing installations and measures for protection against electrical shock* (3rd ed.). IEC.

Lech, M., Skutnik, Z., Bajda, M., & Markowska-Lech, K. (2020). Applications of electrical resistivity surveys in solving selected geotechnical and environmental problems. *Applied Sciences*, *10*(7), 2263. https://doi.org/10.3390/app10072263

Mohamad Nasir, N. A. F., Ab Kadir, M. Z. A., Osman, M., Abd Rahman, M. S., Ungku Amirulddin, U. A., Mohd Nasir, M. S., Zaini, N. H., & Nik Ali, N. H. (2021). Influence of lightning current parameters and earthing system designs on tower footing impedance of 500 kV lines. *Energies*, *14*(16), 4736. https://doi.org/10.3390/en14164736

Mohd Tadza, M. Y., Mohamad, D., Tripathy, S., Rahman, R. A., & Ismail, M. A. M. (2019). Bentonite and marconite for electrical grounding applications from geotechnical engineering perspective. In *AIP Conference Proceedings* (Vol. 2129, No. 1, p. 020078). AIP Publishing. https://doi.org/10.1063/1.5118086

Mostafa, M., Anwar, M. B., & Radwan, A. (2018). Application of electrical resistivity measurement as quality control test for calcareous soil. *HBRC Journal*, *14*(3), 379–384. https://doi.org/10.1016/j.hbrcj.2017.07.001

Patra, S., Sahoo, S., Mishra, P., & Mahapatra, S. C. (2018). Impacts of urbanization on land use/cover changes and its probable implications on local climate and groundwater level. *Journal of Urban Management*, *7*(2), 70–84. https://doi.org/10.1016/j.jum.2018.04.006

Sinchi-Sinchi, F., Coronel-Naranjo, C., Barragán-Escandón, A., & Quizhpi-Palomeque, F. (2022). Soil treatment to reduce grounding resistance by applying low-resistivity material (LRM) implemented in different grounding systems configurations and in soils with different resistivities. *Applied Sciences*, *12*(9), 4788. https://doi.org/10.3390/app12094788

Swileam, G. S., Shahin, R. R., Nasr, H. M., & Essa, K. S. (2019). Spatial variability assessment of Nile alluvial soils using electrical resistivity technique. *Eurasian Journal of Soil Science*, *8*(2), 110–117. https://doi.org/10.18393/ejss.528851

Toh, L. S., & Leong, K. W. (2023). Foundation earthing system – Its application and electrical safety considerations. *IEM Journal*, *83*(1). https://doi.org/10.54552/v83i1.172

Trifunovic, J. (2018). A mathematical method for determining optimal quantity of backfill materials used for grounding resistance reduction. *Mathematical Problems in Engineering*, *2018*, 4863702. https://doi.org/10.1155/2018/4863702

Wan Ahmad, W. F. H., Lai, W. L., Jasni, J., & Ab-Kadir, M. Z. A. (2018). Variations of soil resistivity values due to grounding system installations with natural enhancement material mixtures. In *34th International Conference on Lightning Protection* (pp. 1-5). IEEE. https://doi.org/10.1109/iclp.2018.8503373

Ward, W. O. C., Wilkinson, P. B., Chambers, J. E., Oxby, L. S., & Bai, L. (2014). Distribution-based fuzzy clustering of electrical resistivity tomography images for interface detection. *Geophysical Journal International*, *197*(1), 310–321. https://doi.org/10.1093/gji/ggu006

Zhang, Z., Dan, Y., Zou, J., Liu, G., Gao, C., & Li, Y. (2019). Research on discharging current distribution of grounding electrodes. *IEEE Access*, *7*, 59287–59298. https://doi.org/10.1109/access.2019.2914058

Pertanika J. Sci. & Technol. 33 (6): 2413 - 2439 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Review Article

Unveiling Insights on Scientometric Exploration of Trends in Power Cable Research – A Review

Khairunnisa Norfaeza Mohamed Khairul Fariz[1], Nik Hakimi Nik Ali[1]*, Mohd Iqbal Mohd Noor[2,3], Dalina Johari[1], Azrul Mohd Ariffin[4] and Mohd Azraei Pangah @ Pa'at[5]

[1]School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, Jalan Ilmu 1/1, 40450 Shah Alam, Selangor, Malaysia
[2]Faculty of Business Management, Universiti Teknologi MARA, Cawangan Pahang, Kampus Raub, Felda Krau, 27600 Raub, Pahang, Malaysia
[3]Institute for Biodiversity and Sustainable Development, Universiti Teknologi MARA, Jalan Ilmu 1/1, 40450 Shah Alam, Selangor, Malaysia
[4]College of Engineering / Institute of Power Engineering, Universiti Tenaga Nasional, Jalan Kajang–Puchong, 43000 Kajang, Selangor, Malaysia
[5]High Voltage Cable Diagnostics, Asset Performance Unit, TNB Research Sdn. Bhd., No.1, Lorong Ayer Itam, Kawasan Intitusi Penyelidikan, 43000 Kajang, Selangor, Malaysia

ABSTRACT

Power cables play a crucial role in modern power transmission and distribution networks, serving as vital infrastructure for national grids and power utility companies. However, their high installation, operation, and maintenance costs pose significant challenges, especially as global electricity demand continues to rise. Optimizing power cable technology is essential to ensuring reliable, cost-effective, and sustainable energy systems. These challenges align with the United Nations' Sustainable Development Goals (SDGs), particularly in providing access to affordable and clean energy. While extensive research has been conducted on the theoretical and practical aspects of power cables, a noticeable gap remains in bibliometric and scientometric analyses in this field. Understanding research trends, collaboration networks, and emerging technologies is essential for guiding future developments. This study presents the first comprehensive scientometric analysis of power cable research, addressing the gap by systematically evaluating the field's evolution. Using data from the Web of Science

ARTICLE INFO

Article history:
Received: 29 January 2025
Accepted: 29 April 2025
Published: 25 September 2025

DOI: https://doi.org/10.47836/pjst.33.6.02

E-mail addresses:
2023765459@student.uitm.edu.my (Khairunnisa Norfaeza Mohamed Khairul Fariz)
hakimiali@uitm.edu.my (Nik Hakimi Nik Ali)
mohdiqbalmn@uitm.edu.my (Mohd Iqbal Mohd Noor)
dalinaj@uitm.edu.my (Dalina Johari)
azrula@uniten.edu.my (Azrul Mohd Ariffin)
azraei.pangah@tnb.com.my (Mohd Azraei Pangah @ Pa'at)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

(WoS) database and CiteSpace for visualization, the study examines publication trends, key thematic areas, and the global impact of scholarly contributions. The findings reveal the leading researchers, the most influential publications, the volume of research output by country, and the extent of international collaboration. Additionally, the analysis highlights emerging trends in power cable technology, including advancements in materials, diagnostics, and the integration of smart grids. By mapping the research landscape, this study provides valuable insights for academics, industry professionals, and policymakers to foster collaboration, drive innovation, and enhance the development of sustainable power infrastructure.

Keywords: Cluster analysis, power cables, research analysis, scientometric, Web of Science

INTRODUCTION

Power cables are integral to the functioning of power utility companies and national grids, primarily due to their role in transmitting electrical power. These cables, often made of copper or aluminum conductors with electrical insulation (Powers, 1994), are crucial for maintaining a stable supply of electricity. Notably, cross-linked polyethylene (XLPE) cables are widely used for their mechanical and electrical advantages, including superior heat resistance and flexibility in installation (Illias et al., 2016; Z.-P. Zhang et al., 2020). However, the increasing global demand for electricity and the integration of smart grid systems (M.-F. Chen et al., 2019; R. Huang et al., 2013) present significant challenges in terms of maintaining an uninterrupted energy supply and managing the escalating costs associated with power cable infrastructure. Concerns over energy security compound the growing reliance on power cables (Nick et al., 2016) and the affordability of electricity. This is particularly pertinent considering the United Nations' SDG of ensuring accessible and clean energy for all (Arora & Mishra, 2019). Natural disasters and global events like the COVID-19 pandemic further exacerbate the situation (Carere et al., 2020), which has heightened the need for a robust and sustainable electricity supply, underscores the critical role of power cables in the energy sector.

Despite the evident importance of power cables and extensive research on theoretical and practical aspects since 1994, there has been a lack of scientometric analysis or quantitative view in this field (Xu et al., 2022). Bibliometric analysis has gained popularity in recent years in social science and commercial studies (Donthu et al., 2021). There are also articles focusing on the role of scientific research in advancing sustainable development, particularly through energy and engineering disciplines (Kumar et al., 2023; Raman et al., 2024). Thus, this study aims to fill this gap by employing scientometric methods (Sarquah et al., 2022) to analyze scientific research and publications related to power cables systematically. Specifically, the study aims to address the following research question: i. What are the overall trends in research publications on power cables? ii. What are the central topics or clusters within this research domain? iii. Which publications have been most

influential in these areas? and iv. What are the key publications and keywords that have had a significant impact in the field? This scientometric analysis represents the first of its kind in the power cable industry. It provides a novel perspective by examining the relationships between data, mapping, and clustering networks, analyzing publication output by countries, identifying the most productive authors, and exploring international collaborations (Hwalla et al., 2023). This approach offers a comprehensive understanding of the evolving trends and pivotal elements in power cable research, thereby addressing the identified research questions and contributing valuable insights into this rapidly developing field.

This paper is structured into five main sections: Introduction, Methodology, Results, Discussion, and Conclusion, to provide a comprehensive scientometric analysis of power cable research. The Introduction Section highlights the significance of power cables in the energy industry, as well as their contribution and novelty in addressing the current scientometric study gap. The Methodology Section outlines the approach employed, including database searches conducted within the WoS and the analysis techniques used, which comprise descriptive and scientometric analyses, as well as visualization techniques utilizing CiteSpace. The Results Section presents key findings from the descriptive analysis, including the most cited articles, publication trends per year, publication affiliations, research trends across countries, and author productivity. Meanwhile, the scientometric analysis results include cluster mapping, document co-citation analysis, article burst trends over the years, author co-citation analysis, and keyword burst trends over the years. The Discussion Section synthesizes these findings, linking them to broader technological and policy trends, such as the adoption of renewable energy and advancements in cable insulation and diagnostics. Lastly, this paper concludes by emphasizing the value of scientometric analysis in identifying research gaps and guiding future studies. It also suggests areas for further exploration, including collaborations between leading institutions and comparative analyses using alternative bibliometric tools.

METHODOLOGY

The methodology of this study employs a two-pronged approach, combining both descriptive and scientometric analyses, which draws upon data from the WoS database. The aim of the overarching is to present a comprehensive picture of research trends, influential authors, and key themes within the domain of power cable research.

Database Searches

The foundational step in this study involved a systematic search of the WoS Core Collection database. This is a widely acknowledged resource in scholarly research because of its thorough coverage, excellent citation records, and strict indexing guidelines. The WoS is well known for providing reliable and trustworthy data for bibliometric research,

particularly in ensuring consistency in impact analyses and citation metrics. To preserve the accuracy and dependability of our findings, our paper aimed to prevent potential redundancies and inconsistencies that could arise from using multiple databases.

To ensure a comprehensive search, the study utilized the term "TS" for topic search in WoS, targeting titles, abstracts, keywords, and author details. The search phrase employed was "TS= ("power cable*")," including an asterisk to capture variations of the term. The time frame for the search spanned from 2006 to 2023, with a focus exclusively on original research articles. Exclusions were made for non-research-based publications such as editorials, book chapters, and discussion papers. The search, conducted on December 22, 2023, aimed to create a diverse dataset encompassing a range of research methodologies.

Data Analysis

Data analysis in this study is bifurcated into descriptive analysis and scientometric analysis, detailed as follows:

Descriptive Analysis

The descriptive component focused on quantifying the annual publication output, identifying prominent journals, and pinpointing the most productive authors and institutions. Additionally, it included an analysis of geographical distribution by determining the countries with significant contributions to the field.

Scientometric Analysis

Utilizing CiteSpace 6 (version 6.2 for Windows), the study engaged in a scientometric evaluation through dual map overlay and document co-citation analysis. The dual map overlay method differentiated between cited and citing journals, creating a visual representation of inter-journal relationships. It highlighted the flow of citations, with the size and thickness of map elements reflecting the volume of publications and the frequency of citations, respectively. document co-citation analysis (DCA) was conducted to identify patterns of co-citation among articles. It provided insights into the interconnectedness of publications, revealing key research articles through metrics like burstiness, centrality, and sigma. Burstiness highlighted articles with sudden increases in citations, centrality identified influential papers within the network, and sigma combined these factors to assess the novelty and impact of the research.

For clustering analysis, the study employed a multidimensional clustering approach using the log-likelihood ratio (LLR) for label extraction. Cluster visualization was achieved through a timeline view, illustrating the chronological development of the research field, and a spatial network representation, showcasing the thematic structure of the dataset. Cluster quality was assessed using the modularity index (Q) and the silhouette metric to

evaluate the reliability and coherence of the identified clusters. The methodology's flow and sequence are encapsulated in the framework provided in Figure 1. This structured

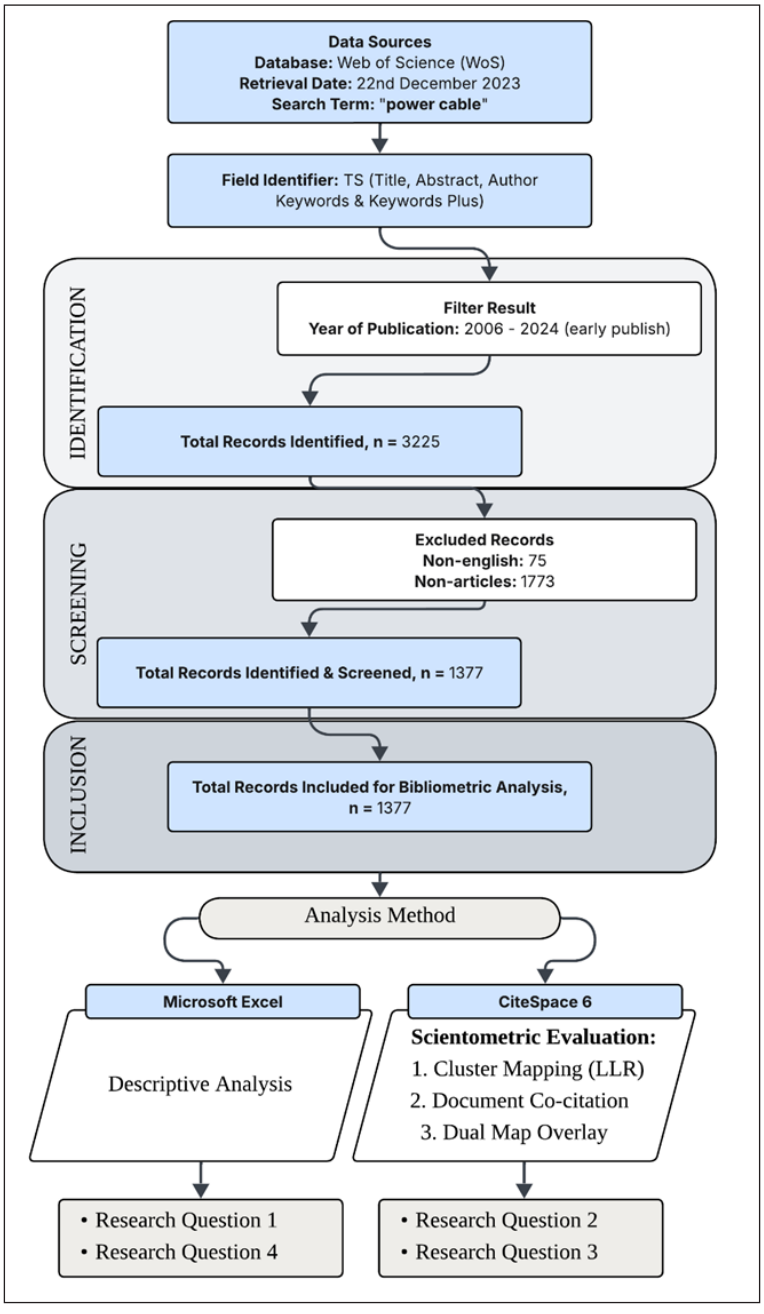


Figure 1. Methodology flow for descriptive and scientometric analysis in this study, consisting of the identification, screening, and inclusion stages
Note. LLR = Log-likelihood ratio

approach ensures a comprehensive and systematic exploration of the power cable research landscape, aiming to uncover patterns, trends, and influential contributors in the field.

RESULTS

Based on the methodology part, the general data gathered has been tabulated as shown in Table 1. As in the methodology section, 1,377 articles retrieved on 22nd December 2023, and published between 2016 and 2024, will be analyzed in this work. The total number of citations recorded is 15,204, and the average number of citations per article is 11.04, which highlights the notable impact of the research over the specified timeframe. Furthermore, the Hirsch index (*h*-index), which measures the number of papers in which other writers have cited a writer, is 52, showing the significance of the potential for co-citations in multidisciplinary domains.

Table 1
General statistics from selected records

Publication titles	Record count
Articles	1,377
Sum of the times cited	15,204
Average citations per item	11.04
h-index	52

This paper presents a scientometric examination of articles published between 1970 and 2024, which serves as the basis for the findings, revealing how references were distributed over time. Older references should also be included to document the changes and patterns that have occurred in the field over the past fifty years. Additionally, since this research was conducted and written at the early stage of data availability from 2024 publications, only one article from 2024 is included. The following section utilizes these general statistics to present the most significant and often cited articles, annual trends in article citations, affiliations, and record trends across the country, as well as notable writers and their record counts.

Most Cited Articles

Table 2 shows the top ten articles (Hammerl et al., 2000; Sun et al., 2012) with their total citations, respectively. From this result, the most influential paper in the power cable research field is cited (Hammerl et al., 2000) with a total of 225 citations, the research demonstrated that calcium doping of grain boundaries in polycrystalline $YBa_2Cu_3O_7$ significantly enhances its supercurrent density at 77 K. Next, the fewest citations among the articles is (Sun et al., 2012), which is 98 citations recorded. These 10 articles, overall, focus on advancements in high-temperature superconductors, electrical insulation, power electronics, fault detection, and biomedical engineering. The key themes include enhancing supercurrent density, optimizing charge injections, improving insulation properties and fault detection methods, and developing efficient power systems for industrial and biomedical applications.

Table 2
Number of citations from the top ten most cited papers

Titles	Total citations
Enhanced supercurrent density in polycrystalline YBa_2Cu_3O(7-delta) at 77 K from calcium doping of grain boundaries (Hammerl et al., 2000)	225
Electrodes and charge injection in low-density polyethylene - Using the pulsed electroacoustic technique (G. Chen et al., 2001)	147
Effect of insulation properties on the field grading of solid dielectric DC cable (Boggs et al., 2001)	143
An efficient method based on the electromagnetic time reversal to locate faults in power networks (Razzaghi et al., 2013)	140
Evaluation of switching performance of SiC devices in PWM inverter-fed induction motor drives (Z. Zhang et al., 2015)	117
Feature extraction of partial discharge signals using the wavelet packet transform and classification with a probabilistic neural network (Evagorou et al., 2010)	113
A distributed optical fibre sensor for temperature detection in power cables (Yilmaz & Karlik, 2006)	109
An integrated driving/charging switched reluctance motor drive using three-phase power module (H.-C. Chang & Liaw, 2011)	100
Electrical treeing characteristics in XLPE power cable insulation in frequency range between 20 and 500 Hz (G. Chen & Tham, 2009)	100
A two-hop wireless power transfer system with an efficiency-enhanced power receiver for motion-free capsule endoscopy inspection (Sun et al., 2012)	98

Article Publication Per Year

As shown in Figure 2, early research publications in the power cable field began to appear in 1970. The number of records did not exceed 20 articles until 2006, indicating that this

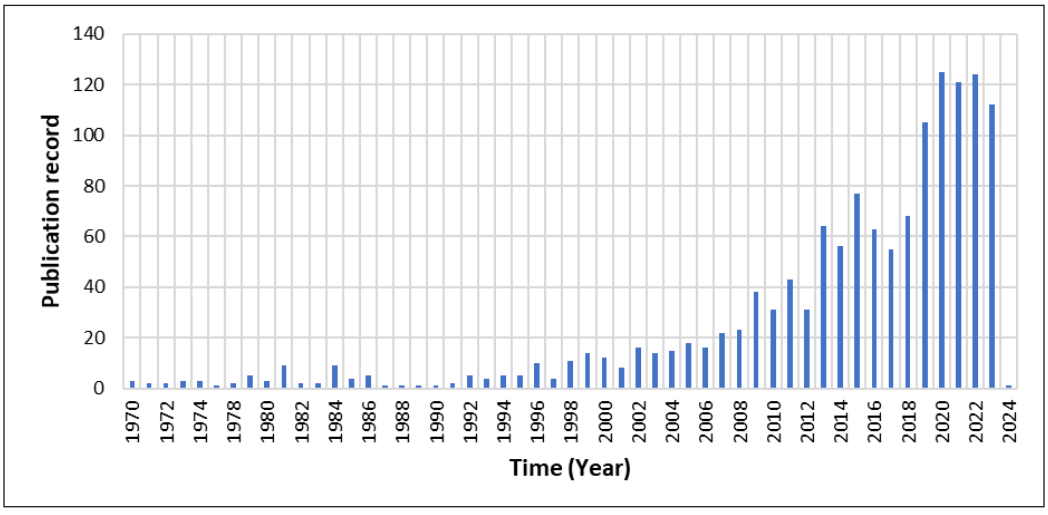


Figure 2. Yearly article publications trend from 1970 to 2024 regarding power cables

study was still in the exploratory stage. However, it was found in an article published in 1970 (Kortschinski & Leslie, 1970). References have been cited from papers published as early as 1931, which observed cable temperatures. Based on that reference, this paper highlights the design features and installations of a thin, liquid-filled coaxial sensing cable for detecting and locating hot spots in underground power cables.

Moving forward to 2013, the number of research publications increased significantly until 2023, indicating that many research studies were conducted widely during this period. Within this period, the year 2020 reached the highest value, which is 125 publications. This increment in number can be observed to be affected by the outbreak of the COVID-19 pandemic that started in 2019, where globally, most activities were carried out online, which then led to unpredictable load changes (Carere et al., 2020; H. Zhong et al., 2020; Zarbakhsh et al., 2022). Besides, the expansion of renewable energy generation and smart grids (Bonnard et al., 2020; Colin & Pilgrim, 2020; Lecuna et al., 2020; Wang et al., 2020), those who rely on intermittent weather conditions also contribute to this number. As one of the essential components of the power system, the power cable became a popular subject during this period. The focus was on enhancing the reliability of power cables to withstand realistic dynamic loads and weather conditions, thereby optimizing conservative cable sizing and rating, and preventing cable failure. Lastly, in 2023, power cables remain a topic of discussion, with 112 publications, indicating that the topic of power cables is still relevant.

Publication Affiliations

Throughout the period of the subject, Table 3 displays the top ten influential affiliations that have contributed to the publication of power cable studies. Affiliations from China have contributed the highest number of publications, followed by Korea, Egypt, and then Italy.

Table 3
Top ten influential affiliations with record numbers

Affiliations	Records count	Country
Xi'an Jiaotong University	**44**	China
State Grid Corporation of China	**36**	China
Changwon National University	**34**	South Korea
Egyptian Knowledge Bank EKB	**31**	Egypt
China Southern Power Grid	**30**	China
Chongqing University	**30**	China
Korea Electrotechnology Research Institute Keri	**29**	Korea
University of Bologna	**29**	Italy
Korea Electric Power Corporation - KEPCO	**27**	Korea
North China Electric Power University	**27**	China

According to the table, Xi'an Jiaotong University in China has the highest productivity in publications, with 44 records. The State Grid Corporation of China, the world's largest power company and a government-owned entity, has 36 records, making it the second highest. Korea's Changwon National University, which currently owns 34 records, came next. The list continues, totaling 317 records. With just 27 records, the two lowest-ranked institutes demonstrate their ongoing improvement. Thus, China appears to be the home of the major research institutes, which have had a significant impact on the field of power cable research.

Records Trend Across Countries

Figure 3 illustrates the global trend in publications. Referring to the blue scale, which ranges from 1 to 363 records, China has the highest number of publications, with 363 records. The United States of America (USA) follows with almost half that number, at around 163 records. Japan and South Korea come next, with 136 and 126 records, respectively. This trend reflects the fact that these countries are among the world's largest economies. The list then continues with countries such as England, Italy, Germany, France, Poland, Canada, Sweden, Taiwan, and Egypt, arranged according to the color concentration on the blue scale.

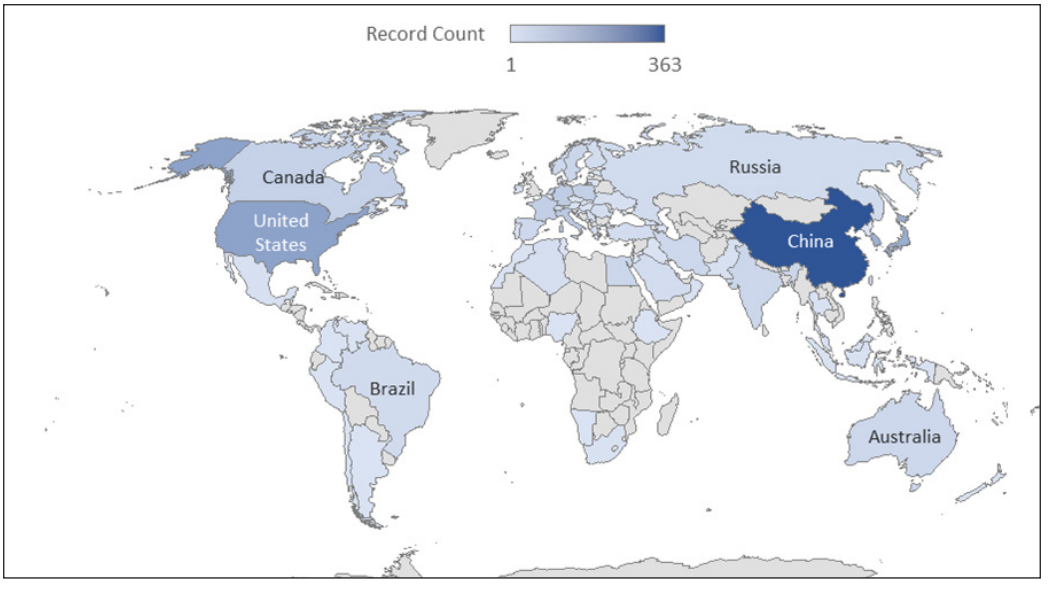


Figure 3. Publications trends across countries are represented by the concentrations on the blue scale, as shown in the legend

Authors' Productivity

As shown in Table 4, the top 10 authors are involved in the areas of electrical power engineering and cable research. For example, Minwon Park, who has 21 publications and

studies in the fields of magnetic fields, superconductors, and critical current, is affiliated with Changwon National University. Next is Peter A. A. F. Wouters, a major in power cables, phase velocity, and dielectric constant, who has published 17 papers with Eindhoven University of Technology. After that, in third and fourth positions, Takato Masuda and Giovanni Mazzanti have each produced 16 publications, specializing in cable systems and thermal stress research. Lastly, the fewest publications are from Dardan Klimenta, a University of Pristina, Kosovo-affiliated author of 11 papers on energy systems, renewable energy sources, and capital costs.

Table 4
Top ten authors with publications retrieved from the Web of Science database

Researcher profiles	Records count	Research areas	Affiliations	Countries
Minwon Park (Park, Minwon)	21	Magnetic field, superconductors, critical current	Changwon National University	South Korea
Peter A. A. F. Wouters (Wouters, P. A. A. F.)	17	Power cables, phase velocity, dielectric constant	Eindhoven University of Technology	Netherlands
Takato Masuda (Masuda, T.)	16	Superconducting cable, cable system, power cable	Sumitomo Electric Industries Ltd.	Japan
Giovanni Mazzanti (Mazzanti, Giovanni)	16	Space charge, cable system, thermal stress	University of Bologna	Italy
In-Keun Yu (Yu, In-Keun)	15	Renewable energy, power system, photovoltaic power	Changwon National University	South Korea
Sastry Pamidi (Pamidi, Sastry)	13	Superconductors, cryogenics, electrical insulation	FAMU-FSU College of Engineering	Florida
Atsushi Ishiyama (Ishiyama, Atsushi)	12	Magnetic field, current flow, superconductors	Waseda University	Japan
Masayoshi Ohya (Ohya, M.)	12	Superconductor, liquid hydrogen, critical current	Sumitomo Electric Industries	Japan
Chul Han Kim (Kim, C. H.)	12	Superconductor, magnetic field, power cables	Florida State University	USA
Dardan Klimenta (Klimenta, Dardan)	11	Energy system, renewable energy, capital costs	University of Pristina	Kosovo

Scientometric Analysis

In this section, the scientometric results, including cluster mapping, references, document burst trend, author co-citations, and keyword trends, are analyzed and briefly described.

CiteSpace, a data visualizer software, is used in this study to visualize or plot the extracted data. Furthermore, CiteSpace can cluster the items identified by optional algorithms such as LLR and evaluate the burst of items to identify keywords with long-term burst and high strength (Liang et al., 2017).

Cluster Mapping Analysis

Figure 4 displays the clusters of records related to the keyword used in this paper. Most academic records were published in top-tier articles in the research area of power cables, including underground transmission lines, underground power cable systems, fault location, material advancements, overcurrent characteristics, and high-temperature applications.

Based on the size of clusters, the top three clusters, Cluster #0, Cluster #1, and Cluster #2, can be considered articles that have a significant influence in this area. They are studied under the topics of underground transmission lines, underground power cable systems, and fault locations, respectively, which are the three key elements in the study of electrical power systems. As shown in Figure 4, the label indicates the item, and the size of clusters is directly proportional to the number of publications in each cluster that included the keyword (Li & Li, 2018). The red spots on each label show the largest cluster in which the most articles were published in relation to their label in the average years.

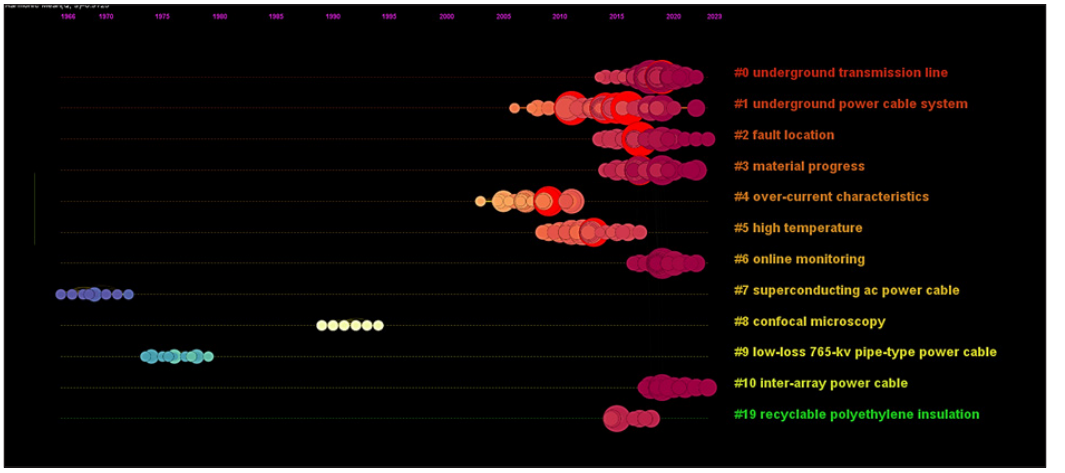


Figure 4. Cluster mapping based on cluster labels

Table 5 is shown to detail the information illustrated by Figure 4. The homogeneity of a cluster is represented by the "silhouette" in the table, the larger the silhouette, the stronger the relationship between the items in the clusters. Cluster #0, with the label “underground transmission line”, is the largest cluster, consisting of 63 articles, and is one of the hotspots in the research area in the 2018 average year. The second cluster, Cluster #1, labeled as

the "underground power cable system", had an average of 50 articles per year in 2014. The next cluster, Cluster #3, has 44 articles published in the average year of 2017. This was followed by the other nine clusters, each with its respective label, as shown in Table 5. Lastly, the top three often cited articles for the top three largest clusters are shown in Tables 6-8.

Table 5
Cluster IDs and details based on cluster label

Cluster ID	Size	Silhouette	Label	Average year
0	63	0.927	Underground transmission line	2018
1	50	0.972	Underground power cable system	2014
2	44	0.976	Fault location	2017
3	43	0.994	Material progress	2017
4	37	1	Over-current characteristics	2007
5	37	0.99	High temperature	2012
6	34	0.975	Online monitoring	2018
7	29	1	Superconducting AC power cable	1968
8	20	1	Confocal microscopy	1990
9	20	1	Low-loss 765 kv pipe-type power cable	1976

Note. AC = Alternating current

Cluster #0**.** In this cluster, the most cited articles are listed in Table 6, with two published in 2018 and one in 2019. The research (Shabani & Vahidi, 2019) introduces an algorithm designed to calculate the most efficient cable ampacity. By utilizing heuristic optimization (ICA) and a probabilistic approach (Monte Carlo), the algorithm's objective function considers cable arrangement and backfill dimensions, incorporating parameter uncertainties to optimize ampacity and load current conformity while minimizing installation costs.

Similarly, the second paper (Cichy et al., 2018) employed a genetic algorithm to address the optimization of the underground power cable conductor. It considered factors such as ampacity, material, and labor costs in cable production, as well as the expenses associated with losses during its operation. Lastly, the next paper (Bustamante, Mínguez, et al., 2019) computed the ampacity of underground cables, incorporating variables such as soil resistivities and depths. The aim was to assess the risk of exceeding the maximum permissible cable temperature based on these variables, which is necessary to determine the cable's longevity.

Cluster #1**.** Table 7 lists the articles that have been the highlights in Cluster #1. Published in 2011, 2014, and 2016, respectively, the publications discuss the thermal analysis and performance analysis of underground electrical power cables. In order to ascertain the thermal resistance that exists between an underground electrical power cable and the ground

Table 6
Most cited articles in Cluster #0

Titles	Year
A probabilistic approach for optimal power cable ampacity computation by considering uncertainty of parameters and economic constraints (Shabani & Vahidi, 2019)	2018
Economic optimization of an underground power cable installation (Cichy et al., 2018)	2018
Thermal behaviour of medium-voltage underground cables under high-load operating conditions (Bustamante, Mínguez, et al., 2019)	2019

surface, the first paper (de Lieto Vollaro et al., 2011) carried out a significant number of numerical simulations for various values of the trench depth and width, the burial depth of the cable, the thicknesses of the two backfilling layers, and the cable bedding, along with their respective thermal conductivities and the thermal conductivity of the mother soil.

Aside from that, a new thermal backfill material, Lafarge Gruntar, is suggested in the following research (Ocłoń et al., 2016). Its performance in relation to water content and time is analyzed for various mass fractions, such as 5%, 10%, and 15%, to maximize the underground power cables' thermal performance. The next paper (Kroener et al., 2014) also investigated the thermal behavior of the underground cables. The heat dissipation from the cables was described by a computational model of linked liquid water, vapor, and heat.

Table 7
Most cited articles in Cluster #1

Titles	Year
Thermal analysis of underground electrical power cables buried in non-homogeneous soils (de Lieto Vollaro et al., 2011)	2011
The performance analysis of a new thermal backfill material for underground power cable (Ocłoń et al., 2016)	2016
Numerical simulation of coupled heat, liquid water and water vapor in soils for heat dissipation of underground electrical power cables (Kroener et al., 2014)	2014

Cluster #2**.** In this cluster, Cluster #2 has three other articles published, as shown in Table 8, that discuss cable durability and lifetime. To maximize the operational value of cable assets while maintaining the cable life cycle, a 2017 paper (Zhou et al., 2017) offers a thorough analysis of current research and development in the areas of cable fault diagnosis, condition monitoring and diagnosis, life assessment techniques, fault location, and optimization of maintenance and replacement strategies.

Finally, the following two studies (C.-K. Chang et al., 2019; Dong et al., 2017) demonstrated the simulation of sheath system currents in a cross-bonded cable system analysis and offered affordable and workable decision tree rules for an online partial discharge (PD) monitoring device to assess the viability of fault diagnosis and accomplish condition-based maintenance in distribution power cables.

Table 8
Most cited articles in Cluster #2

Titles	Year
Review of recent research towards power cable life cycle management (Zhou et al., 2017)	2017
Online monitoring and diagnosis of HV cable faults by sheath system currents (Dong et al., 2017)	2017
Decision tree rules for insulation condition assessment of pre-molded power cable joints with artificial defects (C.-K. Chang et al., 2019)	2019

References Analysis (Document Co-citation Analysis)

Researchers might access several of the most relevant articles with creative concepts and trustworthy experimental results. Thus, this document's co-citation analysis is performed to assist in identifying important literature for cross-disciplinary concepts. Figure 5 illustrates the interaction between high-citation articles. The node's label represents the name and publication year of the author, and the thickness of the node is determined by the frequency of its mention.

According to Figure 5, studies by Bicen (2017), Bustamante, Manana, et al. (2019), and Dinmohammadi et al. (2019), are among the highly cited articles based on the node size. Briefly, their papers highlight the monitoring of underground power cable lifetimes using a trend-adjusted methodology, calculate dissolved gas analysis (DGA) limit values and optimal sampling intervals, and develop predictive models for assessing damage and

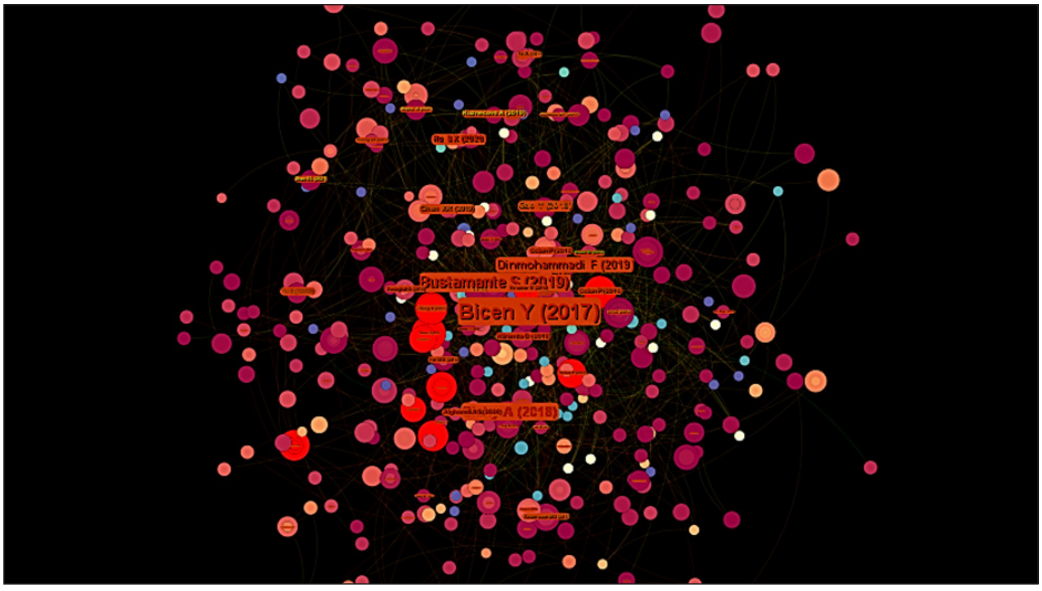


Figure 5. Articles' co-citation analysis on power cables research

estimating the life expectancy of subsea power cables, aiming to enhance reliability and inform maintenance planning.

To provide better clarity for analysis, Table 9 shows the centrality and sigma values of each article that cited authors' articles (displayed in Figure 5). The centrality of 0.01 for each article means that the same factors have the same values in terms of how close they are to one another and how information flows through them. Sigma, on the other hand, is the total of the burstiness and centrality scores, with the highest value corresponding to high-value research publications. Therefore, the top three articles, with a sigma value of 1.04, are the most influential ones, followed by the two articles with a sigma score of 1.03.

Collectively, the research in Table 9 encompasses advancements in thermal management and diagnostic techniques for underground power cables. Studies on novel thermal backfill materials and coupled heat, liquid water, and vapor movements in soils demonstrate significant improvements in heat dissipation and cable performance. The reviews of power cable life cycle management highlight emerging diagnostics and maintenance strategies enhancing reliability. Next, innovations in online monitoring using sheath system currents provide real-time fault detection for high-voltage cables. Lastly, numerical simulations of heat dissipation processes in thermal backfills and multilayered soils underscore the importance of material and design choices in optimizing underground cable efficiency and lifespan.

Table 9
The top five co-citation articles score

Title	Centrality	Sigma
Numerical simulation of coupled heat, liquid water and water vapor in soils for heat dissipation of underground electrical power cables (Kroener et al., 2014)	0.01	1.04
The performance analysis of a new thermal backfill material for underground power cable system (Ocłoń et al., 2016)	0.01	1.04
Review of recent research towards power cable life cycle management (Zhou et al., 2017)	0.01	1.04
Numerical simulation of heat dissipation processes in underground power cable system situated in thermal backfill and buried in a multilayered soil (Ocłoń et al., 2015)	0.01	1.03
Online monitoring and diagnosis of HV cable faults by sheath system currents (Dong et al., 2017)	0.01	1.03

Note. HV = High-voltage

Document Burst Trend Over the Years

Table 10 depicts the trend of top publications with the most powerful citation bursts, where the period of each burst is represented by the red line on the blue timeline (from 1970 to

Table 10
The burst trend of the most influential documents between 1970 and 2023

References	Year	Strength	Begin	End	1970 - 2023
A new HTS cable project in Japan (Masuda et al., 2009)	2009	5.49	2011	2013	
Phase II of the Albany HTS cable project (Yumura et al., 2009)	2009	4.26	2011	2013	
Thermal analysis of underground electrical power cables buried in non-homogeneous soils (de Lieto Vollaro et al., 2011)	2011	5.85	2013	2016	
Numerical simulation of coupled heat, liquid water, and water vapor in soils for heat dissipation of underground electrical power cables (Kroener et al., 2014)	2014	5.26	2015	2018	
Update of YOKOHAMA HTS cable project (Yumura et al., 2013)	2013	4.54	2015	2017	
Numerical simulation of heat dissipation processes in underground power cable system situated in thermal backfill and buried in a multilayered soil (Ocłoń et al., 2015)	2015	5.85	2016	2018	
The performance analysis of a new thermal backfill material for underground power cable system (Ocłoń et al., 2016)	2016	4.55	2017	2021	
Linescout Technology opens the way to robotic inspection and maintenance of high-voltage power lines (Pouliot et al., 2015)	2015	4.74	2018	2019	
Review of recent research towards power cable life cycle management (Zhou et al., 2017)	2017	4.82	2019	2023	
Online monitoring and diagnosis of HV cable faults by sheath system currents (Dong et al., 2017)	2017	4.42	2019	2023	
A probabilistic approach for optimal power cable ampacity computation by considering uncertainty of parameters and economic constraints (Shabani & Vahidi, 2019)	2019	5.57	2020	2023	
Polypropylene based thermoplastic polymers for potential recyclable HVDC cable insulation applications (X. Huang et al., 2017)	2017	4.17	2020	2023	

Note. HTS = High-temperature superconducting; HVDC = High voltage direct current

2023). A burst reflects the emergence of a keyword in a publication during a specific time. As some of the articles from this table have been reviewed in the previous subsection, the remaining articles (Masuda et al., 2009; X. Huang et al., 2017) will be reviewed in this subsection first, and then the burst trend will be explained.

The articles focus on advancements in high-temperature superconducting (HTS) and high-voltage power cable technologies. Masuda et al. (2009) discuss a new HTS cable project in Japan aimed at enhancing power transmission efficiency using superconductors. Yumura et al. (2009) cover Phase II of the Albany HTS Cable Project, emphasizing improvements in cable performance and reliability. Their 2013 update (Yumura et al., 2013) on the Yokohama HTS Cable Project highlights progress in the deployment and operational testing of HTS cables.

Next, Pouliot et al. (2015) introduce LineScout Technology, a robotic system for inspecting and maintaining high-voltage power lines, aiming to increase safety and efficiency in power line maintenance. Lastly, X. Huang et al. (2017) explore polypropylene-based thermoplastic polymers for HVDC cable insulation, focusing on the potential for recyclability and improved electrical properties. Together, these studies aim to enhance power transmission infrastructure through innovative materials and technologies.

From the table, the third (de Lieto Vollaro et al., 2011) and sixth (Ocłoń et al., 2015) publications that studied the underground power cable thermal analysis and heat dissipation have the same strongest burst, 5.85, which were published in 2011 and 2015, respectively. The papers were maintained as popular publications for four years, from 2013 to 2016, and for three years, from 2016 to 2018, respectively. Next, the second last paper that was published in 2019 (Shabani & Vahidi, 2019) is one of the highest strengths of burst, 5.57. This paper started to call out a year after its publication, that is, 2020, and remained the hot topic in power cable research until 2023, along with the last paper that was published in 2017 (X. Huang et al., 2017) with the strength of the burst is 4.27. The other two papers that are still bursting until 2023 are Zhou et al. (2017) and Dong et al. (2017), which were both published in 2017 with burst strengths of 4.48 and 4.42.

Authors Co-citation Analysis

The network of authors' co-citations is shown in Figure 6, with the thickness of the lines connecting authors' names signifying the degree of collaboration. The font size corresponds to the number of published articles on which the other authors have collaborated, and the publication date determines the color of the ring. Since the author's name stood out the most among the others, it is evident that R. Bartnikas et al. is the most closely linked author. Next, G. Mazzanti came in second, followed by A. Ametani in third.

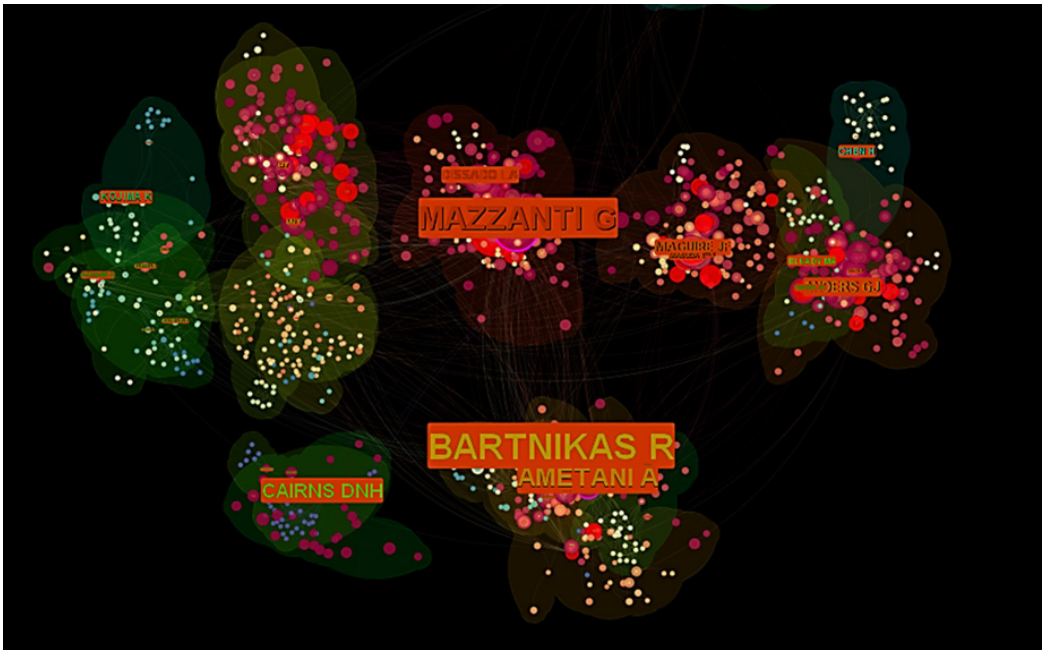


Figure 6. Networks of authors' co-citations

Nevertheless, the authors listed in Table 11 below are the most pertinent, as they are the top co-citation authors related to this field after narrowing the search parameters. James F. Maguire came in first with 0.07 centrality and a 1.84 sigma value. Given that J. F. Maguire's papers have the greatest sigma value, which presents a high degree of centrality, this indicates that he is a highly valued author of research articles.

Although Gian Carlo Montanari placed second with a 1.67 sigma value, Gian Carlo Montanari has the highest centrality, 0.09, which has a significant impact on the behavior of the research area since the centrality indicates the degree of influence one author has over another. Next, Takato Masuda, who also has a 1.67 sigma value, has a centrality of 0.04. Steven A. Boggs followed this with 0.04 and 1.44 for centrality and sigma value. Finally, Shinichi Mukoyama has the lowest centrality and sigma values, which are 0.02 and 1.32.

Table 11
The authors with the highest score of co-citation

Author	Centrality	Sigma
James F. Maguire	0.07	1.84
Gian Carlo Montanari	0.09	1.67
Takato Masuda	0.04	1.67
Steven A. Boggs	0.07	1.44
Shinichi Mukoyama	0.02	1.32

Keyword Burst Trend Over the Years

The top 20 related keywords with the strongest citation bursts are displayed in Table 12, which displays the concentrations in study areas and new trends over the years. In this keyword analysis, the keywords that the burst line represents an outburst in each field of study at a given time. The time span of each keyword burst is indicated by the corresponding red line on the blue timeline (from 1970 to 2023), as shown in the burst analysis above. From the table, the term "XLPE", which stands for cross-linked polyethylene, a thermoset insulation material commonly used in electrical cable for power transmission and distribution, has the longest burst duration among others.

From 1999 to 2013, the term was used widely for 14 years, with a burst strength of 4.4. Moreover, in 2016, the same keyword, "XLPE insulation", was used and had the second strongest burst strength, which is 11.15, for two years, beginning from 2018 to 2020. The term "power cable insulation" has been in use for 4 years, from 2019 to 2023, and has the highest burst strength, which is 13.3, indicating that it has been widely used in many research documents to date. Besides that, the term "power cables" holds the third-highest burst strength, which is 11.05, and has also become the central topic of research until 2023, along with the terms "cable shielding", "temperature measurement", and "diagnosis", with the burst strengths of 5.77, 4.3, and 4.19, respectively.

DISCUSSION

In this section, the results obtained from the scientometric analysis performed in this paper are briefly described. This discussion answers the research questions stated in the introduction above.

The Overall Publication Trends in Terms of Output

From a general analysis, the fact that the total number of citations from 1,377 retrieved articles was 15,204, and the average number of citations per article was 11.04, indicates the significant impact of the research over the period. The most cited article, which has made the most impactful contributions to the field of power cable research, was identified with 225 citations. Furthermore, from 2013 to 2023, the total number of articles published has grown dramatically, and this trend is expected to continue in the coming years. The results section discusses these trends, highlighting factors like COVID-19 and the integration of renewable energy and smart grids. These are driven by rapid technological advancements, the urgency to address climate change, and policies, subsidies, and incentives promoting renewable energy and smart grids to achieve SDGs such as SDG 7, SDG 9, SDG 11, SDG 12, and SDG 13 (Arora & Mishra, 2019).

Table 12
The burst trend of the notable keywords from 1970 to 2023

Keywords	Year	Strength	Begin	End	1970 - 2023
XLPE	1999	4.40	1999	2013	
Dielectric breakdown	2001	5.86	2001	2009	
AC loss	2004	6.56	2004	2013	
Superconducting cables	1992	7.33	2006	2011	
Superconducting power cable	2007	5.11	2007	2017	
Transmission lines	2008	5.24	2008	2015	
System	2008	4.23	2008	2018	
High-temperature superconductors	2009	5.93	2009	2016	
HTS power cable	2002	4.94	2009	2014	
Fault current	2009	4.52	2009	2016	
Power cable connecting	2009	4.28	2009	2014	
Partial discharge	1996	4.58	2012	2014	
Transients	2013	4.49	2013	2017	
Space charge	2001	4.97	2017	2020	
XLPE insulation	2016	11.15	2018	2020	
Power cable insulation	2011	13.30	2019	2023	
Cable shielding	2009	5.77	2020	2023	
Temperature measurement	2020	4.30	2020	2023	
Power cables	1992	11.05	2021	2023	
Diagnosis	2018	4.19	2021	2023	

Note. XLPE = Cross-linked polyethylene; AC = Alternating current; HTS = High-temperature superconducting

Next, China is the leading country in the field of research on affiliated power cables, as evidenced by the contributions of five institutions to research publications. This is due to the Chinese government's implementation of strong policy initiatives and substantial funding to support research and development (R&D) in the energy sector (Sandalow et al., 2022). However, the USA has published the most resourceful records in the world, followed by Japan and South Korea. This statement is then supported by the fact that the top ten authors are mainly from Japan, Korea, and the USA, having published the most relevant records, totaling 101 publications. This indicates that advanced-economy countries, such as the USA, Japan, and Korea, are at the forefront of this industry, as these countries have dedicated significant government and industry funding to energy research, fostering collaboration on low-carbon and carbon-free energy technologies (U.S. Department of Commerce, 2024).

The Central Topics and Their Temporal Evolution

According to the cluster analysis, "underground transmission line", "underground power cable system", "fault site", and other terms were the most popular study subjects and directions. The focus of research has been on improving cable ampacity and its durability against critical environmental and load conditions.

Moreover, in the past decade, power cable research has focused on improving superconducting cables and power (2010-2017), enhancing insulation (2019-2023), and advancing diagnostic and measurement technologies (2020-2023). Efforts in superconducting materials aimed to increase efficiency and capacity, while insulation research focused on durability and performance. Diagnostic advancements are geared towards predictive maintenance and real-time monitoring. High-temperature superconductors (2016-2023) also received attention for their potential to overcome conventional limitations.

Looking ahead, research will likely concentrate on diagnostic systems, high-performance insulation materials, renewable energy integration, and smart grid technologies soon. In the far future, advancements in superconducting technologies, nanotechnology, eco-friendly materials, and cyber-physical systems will be crucial. Dynamic Line Rating (DLR) or Dynamic Current Rating (DCR) are recent instances of technologies that have been acknowledged as a continuing trend (Fariz et al., 2024). In addition to these advancements, future research is expected to focus on the integration of sensors, artificial intelligence (AI), and machine learning (ML) for real-time grid optimization, fault detection, and predictive maintenance of power transmission systems, as in a study by Radzi et al. (2021), where an ultrasonic sensor is used for detecting partial discharges.

Moreover, the application of advanced thermal modelling techniques, including computational fluid dynamics (CFD) simulations, will improve the accuracy of cable rating assessments, particularly for cables installed in tunnels or densely populated urban

areas. Research into high-voltage direct current (HVDC) transmission and flexible AC transmission systems (FACTS) will also gain momentum, as these technologies provide enhanced efficiency and stability for long-distance power delivery. These evolutions reflect the ongoing drive for efficient, reliable, and sustainable power transmission solutions. Research on these targeted topics has also been ongoing for a substantial period, as the results of this study could have long-term benefits for the national grid and power utility companies, including lower installation, maintenance, and operating costs, while also contributing to the achievement of the SDGs.

The Most Influential Publications in These Domains

The analysis of document co-citation reveals that articles by Bicen (2017), Bustamante et al., and Dinmohammadi et al. (2019) are among the most cited, indicating that their publications have made significant contributions to various studies related to power cables. After the scope of data has been constrained, the most relevant and influential publications are those that study heat and thermal analysis, which influence the cables' lifecycle and cable fault diagnosis. These publications have recently become the focus of interest in the power cables domain. Additionally, since James F. Maguire and Gian Carlo Montanari have a greater degree of influence than other authors, based on their centrality and sigma values in author co-citation analysis, their publications have a substantial impact on the behavior of the research area.

The Most Impactful Publications and Keywords

The concentrations in power cable research areas and new trends over the years have been discovered using document and keyword burst trend analysis. From the results, both articles, which have the strongest burst in document burst analysis, are the most impactful publications, providing clear insight into the thermal and heat dissipation analysis of underground power cables with variable soil parameters. These studies may have catalyzed numerous existing research projects or further potential publications on the thermal behavior of underground power cables.

Apart from that, the most impactful keyword, as indicated by the keyword burst trend analysis, is power cable insulation, a crucial component in the design of power cables that is necessary for the reliable transfer of energy under various load conditions. This keyword may have been applied in conjunction with other keywords mentioned in the results section, as it refers to a common component that prevents heat, sound, or electric current from escaping or entering the cables.

CONCLUSION

Power cables are among the crucial parts, particularly in the field of electrical engineering, that require extensive research to ensure the most dependable and economical installation, maintenance, and use of cables, given their importance as a national grid and electrical utility company asset. In this paper, the scientometric analysis is performed to identify significant research gaps and suggest future paths by assembling a collection of relevant studies from different power cable studies.

Thus, this study has provided a thorough overview of the existing research collection, illustrating the chronological evolution of power cables over time, as well as the influential papers, countries, and authors recognized in this discipline. Furthermore, the core content of this paper is how to use an appropriate search scheme to develop cable from a literature perspective. For future studies, authors could map these research topics and possibly pinpoint missing relationships by comparing other databases with WoS. Subsequent studies may also examine the significant partnerships between nations and writers, as well as compare outcomes utilizing the widely used scientometric analytic software, such as VOSviewer.

ACKNOWLEDGMENTS

The research was supported by funds provided by the Ministry of Higher Education (MOHE) under the Fundamental Research Grant Scheme (FRGS) (FRGS/1/2023/TK07/UITM/02/22) and by the College of Engineering, Universiti Teknologi MARA, Shah Alam.

REFERENCES

Arora, N. K., & Mishra, I. (2019). United Nations Sustainable Development Goals 2030 and environmental sustainability: Race against time. *Environmental Sustainability*, *2*, 339–342. https://doi.org/10.1007/s42398-019-00092-y

Bicen, Y. (2017). Trend adjusted lifetime monitoring of underground power cable. *Electric Power Systems Research*, *143*, 189–196. https://doi.org/10.1016/j.epsr.2016.10.045

Boggs, S., Damon, D. H., Hjerrild, J., Holboll, J. T., & Henriksen, M. (2001). Effect of insulation properties on the field grading of solid dielectric DC cable. *IEEE Transactions on Power Delivery*, *16*(4), 456–461. https://doi.org/10.1109/61.956720

Bonnard, C.-H., Blavette, A., Bourguet, S., & Charmetant, A. (2020). Modeling of a wave farm export cable for electro-thermal sizing studies. *Renewable Energy*, *147*(Part 1), 2387–2398. https://doi.org/10.1016/j.renene.2019.09.135

Bustamante, S., Manana, M., Arroyo, A., Martinez, R., González, A., & Rodriguez, J. I. (2019). Case study - Calculation of DGA limit values and sampling interval in power transformers. In *6th International Advanced Research Workshop on Transformers* (pp. 64–68). IEEE. https://doi.org/10.23919/ARWtr.2019.8930184

Bustamante, S., Mínguez, R., Arroyo, A., Manana, M., Laso, A., Castro, P., & Martinez, R. (2019). Thermal behaviour of medium-voltage underground cables under high-load operating conditions. *Applied Thermal Engineering*, *156*, 444–452. https://doi.org/10.1016/j.applthermaleng.2019.04.083

Carere, F., Bragatto, T., & Santori, F. (2020). A distribution network during the 2020 COVID-19 Pandemic. In *AEIT International Annual Conference* (pp. 1–6). IEEE. https://doi.org/10.23919/AEIT50178.2020.9241191

Chang, C.-K., Lai, C.-S., & Wu, R.-N. (2019). Decision tree rules for insulation condition assessment of pre-molded power cable joints with artificial defects. *IEEE Transactions on Dielectrics and Electrical Insulation*, *26*(5), 1636–1644. https://doi.org/10.1109/TDEI.2019.008208

Chang, H.-C., & Liaw, C.-M. (2011). An integrated driving/charging switched reluctance motor drive using three-phase power module. *IEEE Transactions on Industrial Electronics*, *58*(5), 1763–1775. https://doi.org/10.1109/TIE.2010.2051938

Chen, G., Tay, T. Y. G., Davies, A. E., Tanaka, Y., & Takada, T. (2001). Electrodes and charge injection in low-density polyethylene using the pulsed electroacoustic technique. *IEEE Transactions on Dielectrics and Electrical Insulation*, *8*(6), 867–873. https://doi.org/10.1109/94.971439

Chen, G., & Tham, C. H. (2009). Electrical treeing characteristics in XLPE power cable insulation in frequency range between 20 and 500 Hz. *IEEE Transactions on Dielectrics and Electrical Insulation*, *16*(1), 179–188. https://doi.org/10.1109/TDEI.2009.4784566

Chen, M.-F., Chiu, H.-C., Tseng, K.-S., Yang, Y.-C., Chou, C.-Y., & Jiang, J.-A. (2019). An IoT-based temperature monitoring system for underground cable tunnels. In *IEEE PES GTD Grand International Conference and Exposition Asia* (pp. 960–964). IEEE. https://doi.org/10.1109/GTDAsia.2019.8715849

Cichy, A., Sakowicz, B., & Kaminski, M. (2018). Economic optimization of an underground power cable installation. *IEEE Transactions on Power Delivery*, *33*(3), 1124–1133. https://doi.org/10.1109/TPWRD.2017.2728702

Colin, M. A. H., & Pilgrim, J. A. (2020). Cable thermal risk estimation for overplanted wind farms. *IEEE Transactions on Power Delivery*, *35*(2), 609–617. https://doi.org/10.1109/TPWRD.2019.2917789

de Lieto Vollaro, R., Fontana, L., & Vallati, A. (2011). Thermal analysis of underground electrical power cables buried in non-homogeneous soils. *Applied Thermal Engineering*, *31*(5), 772–778. https://doi.org/10.1016/J.APPLTHERMALENG.2010.10.024

Dinmohammadi, F., Flynn, D., Bailey, C., Pecht, M., Yin, C., Rajaguru, P., & Robu, V. (2019). Predicting damage and life expectancy of subsea power cables in offshore renewable energy applications. *IEEE Access*, *7*, 54658–54669. https://doi.org/10.1109/ACCESS.2019.2911260

Dong, X., Yang, Y., Zhou, C., & Hepburn, D. M. (2017). Online monitoring and diagnosis of HV cable faults by sheath system currents. *IEEE Transactions on Power Delivery*, *32*(5), 2281–2290. https://doi.org/10.1109/TPWRD.2017.2665818

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Evagorou, D., Kyprianou, A., Lewin, P. L., Stavrou, A., Efthymiou, V., Metaxas, A. C., & Georghiou, G. E. (2010). Feature extraction of partial discharge signals using the wavelet packet transform and classification with a probabilistic neural network. *IET Science, Measurement and Technology*, *4*(3), 177–192. https://doi.org/10.1049/iet-smt.2009.0023

Fariz, K. N. M. K., Nik Ali, N. H., Johari, D., Mohd Noor, M. I., Mohd Ariffin, A., & Mohammad Noor, S. Z. (2024). The evolution of dynamic line rating: technological advancements and future prospects. In *14th International Conference on System Engineering and Technology* (pp. 156–161). IEEE. https://doi.org/10.1109/ICSET63729.2024.10774680

Hammerl, G., Schmehl, A., Schulz, R. R., Goetz, B., Bielefeldt, H., Schneider, C. W., Hilgenkamp, H., & Mannhart, J. (2000). Enhanced supercurrent density in polycrystalline $YBa_2Cu_3O_{7-\delta}$ at 77 K from calcium doping of grain boundaries. *Nature*, *407*, 162–164. https://doi.org/10.1038/35025014

Huang, R., Pilgrim, J. A., Lewin, P. L., & Payne, D. (2013). Dynamic cable ratings for smarter grids. In *4th IEEE/PES Innovative Smart Grid Technologies Europe* (pp. 1-5). IEEE. https://doi.org/10.1109/ISGTEurope.2013.6695230

Huang, X., Fan, Y., Zhang, J., & Jiang, P. (2017). Polypropylene based thermoplastic polymers for potential recyclable HVDC cable insulation applications. *IEEE Transactions on Dielectrics and Electrical Insulation*, *24*(3), 1446–1456. https://doi.org/10.1109/TDEI.2017.006230

Hwalla, J., Bawab, J., El-Hassan, H., Abu Obaida, F., & El-Maaddawy, T. (2023). Scientometric analysis of global research on the utilization of geopolymer composites in construction applications. *Sustainability*, *15*(14), 11340. https://doi.org/10.3390/su151411340

Illias, H. A., Tunio, M. A., Bakar, A. H. A., Mokhlis, H., & Chen, G. (2016). Partial discharge phenomena within an artificial void in cable insulation geometry: Experimental validation and simulation. *IEEE Transactions on Dielectrics and Electrical Insulation*, *23*(1), 451–459. https://doi.org/10.1109/TDEI.2015.005155

Kortschinski, J., & Leslie, J. R. (1970). A power-cable temperature monitoring system. *IEEE Transactions on Power Apparatus and Systems*, *PAS-89*(7), 1429–1433. https://doi.org/10.1109/TPAS.1970.292573

Kroener, E., Vallati, A., & Bittelli, M. (2014). Numerical simulation of coupled heat, liquid water and water vapor in soils for heat dissipation of underground electrical power cables. *Applied Thermal Engineering*, *70*(1), 510–523. https://doi.org/10.1016/j.applthermaleng.2014.05.033

Kumar, C., Pattnaik, D., Balas, V. E., & Raman, R. (2023). Comprehensive scientometric analysis and longitudinal SDG mapping of Quality and Reliability Engineering International journal. *Journal of Scientometric Research*, *12*(3), 558–569. https://doi.org/10.5530/JSCIRES.12.3.053

Lecuna, R., Castro, P., Manana, M., Laso, A., Domingo, R., Arroyo, A., & Martinez, R. (2020). Non-contact temperature measurement method for dynamic rating of overhead power lines. *Electric Power Systems Research*, *185*, 106392. https://doi.org/10.1016/j.epsr.2020.106392

Li, X., & Li, H. (2018). A visual analysis of research on information security risk by using CiteSpace. *IEEE Access*, *6*, 63243–63257. https://doi.org/10.1109/ACCESS.2018.2873696

Liang, Y.-D., Li, Y., Zhao, J., Wang, X.-Y., Zhu, H.-Z., & Chen, X.-H. (2017). Study of acupuncture for low back pain in recent 20 years: A bibliometric analysis via CiteSpace. *Journal of Pain Research*, *10*, 951–964. https://doi.org/10.2147/JPR.S132808

Masuda, T., Yumura, H., Ohya, M., Kikuta, T., Hirose, M., Honjo, S., Mimura, T., Kito, Y., Yamamoto, K., Ikeuchi, M., & Ohno, R. (2009). A new HTS cable project in Japan. *IEEE Transactions on Applied Superconductivity*, *19*(3), 1735–1739. https://doi.org/10.1109/TASC.2009.2018157

Nick, M., Alizadeh-Mousavi, O., Cherkaoui, R., & Paolone, M. (2016). Security constrained unit commitment with dynamic thermal line rating. *IEEE Transactions on Power Systems*, *31*(3), 2014–2025. https://doi.org/10.1109/TPWRS.2015.2445826

Ocłoń, P., Bittelli, M., Cisek, P., Kroener, E., Pilarczyk, M., Taler, D., Rao, R. V., & Vallati, A. (2016). The performance analysis of a new thermal backfill material for underground power cable system. *Applied Thermal Engineering*, *108*, 233–250. https://doi.org/10.1016/j.applthermaleng.2016.07.102

Ocłoń, P., Cisek, P., Pilarczyk, M., & Taler, D. (2015). Numerical simulation of heat dissipation processes in underground power cable system situated in thermal backfill and buried in a multilayered soil. *Energy Conversion and Management*, *95*, 352–370. https://doi.org/10.1016/J.ENCONMAN.2015.01.092

Pouliot, N., Richard, P.-L., & Montambault, S. (2015). LineScout Technology opens the way to robotic inspection and maintenance of high-voltage power lines. *IEEE Power and Energy Technology Systems Journal*, *2*(1), 1–11. https://doi.org/10.1109/JPETS.2015.2395388

Powers, W. F. (1994). The basics of power cable. In *Proceedings of 1994 IEEE 36th Cement Industry Technical Conference* (pp. 37–45). IEEE. https://doi.org/10.1109/CITCON.1994.343441

Radzi, M. I. F. M., Ahmad, M. R., Ali, N. H. N., & Ariffin, A. M. (2021). Development of partial discharge detection device for PD detection on medium voltage XLPE cable. In *8th International Conference on Intelligent and Advanced Systems* (pp. 1–5). IEEE. https://doi.org/10.1109/ICIAS49414.2021.9642630

Raman, R., Pattnaik, D., Kumar, C., & Nedungadi, P. (2024). Advancing sustainable energy systems: A decade of SETA research contribution to sustainable development goals. *Sustainable Energy Technologies and Assessments*, *71*, 103978. https://doi.org/10.1016/j.seta.2024.103978

Razzaghi, R., Lugrin, G., Manesh, H. M., Romero, C., Paolone, M., & Rachidi, F. (2013). An efficient method based on the electromagnetic time reversal to locate faults in power networks. *IEEE Transactions on Power Delivery*, *28*(3), 1663–1673. https://doi.org/10.1109/TPWRD.2013.2251911

Sandalow, D., Meidan, M., Andrews-Speed, P., Hove, A., Qiu, S., & Downie, E. (2022). *Guide to Chinese climate policy 2022*. Oxford Institute of Energy Studies.

Sarquah, K., Narra, S., Beck, G., Awafo, E. A., & Antwi, E. (2022). Bibliometric analysis; Characteristics and trends of refuse derived fuel research. *Sustainability*, *14*(4), 1994. https://doi.org/10.3390/su14041994

Shabani, H., & Vahidi, B. (2019). A probabilistic approach for optimal power cable ampacity computation by considering uncertainty of parameters and economic constraints. *International Journal of Electrical Power and Energy Systems*, *106*, 432–443. https://doi.org/10.1016/j.ijepes.2018.10.030

Sun, T., Xie, X., Li, G., Gu, Y., Deng, Y., & Wang, Z. (2012). A two-hop wireless power transfer system with an efficiency-enhanced power receiver for motion-free capsule endoscopy inspection. *IEEE Transactions on Biomedical Engineering*, *59*(11), 3247–3254. https://doi.org/10.1109/TBME.2012.2206809

U.S. Department of Commerce. (2024, June 26). *Joint Statement: Japan-Republic of Korea-United States Commerce and Industry Ministerial Meeting*. https://www.meti.go.jp/press/2024/06/20240627005/20240627005-a.pdf

Wang, L., Kuan, B.-L., Yu, C.-H., Wu, H.-Y., Zeng, S.-Y., Prokhorov, A. V., Mokhlis, H., & Huat, C. K. (2020). Effects of submarine-cable types and parameters on performance of a future-scheduled offshore wind farm connected to Taiwan power system. *IEEE Transactions on Industry Applications*, *56*(2), 1171–1179. https://doi.org/10.1109/TIA.2020.2966159

Xu, J., Lv, T., Hou, X., Deng, X., & Liu, F. (2022). A bibliometric analysis of power system planning research during 1971–2020. *IEEE Transactions on Power Systems*, *37*(3), 2283–2296. https://doi.org/10.1109/TPWRS.2021.3115159

Yilmaz, G., & Karlik, S. E. (2006). A distributed optical fiber sensor for temperature detection in power cables. *Sensors and Actuators A: Physical*, *125*(2), 148–155. https://doi.org/10.1016/j.sna.2005.06.024

Yumura, H., Ashibe, Y., Itoh, H., Ohya, M., Watanabe, M., Masuda, T., & Weber, C. S. (2009). Phase II of the Albany HTS cable project. *IEEE Transactions on Applied Superconductivity*, *19*(3), 1698–1701. https://doi.org/10.1109/TASC.2009.2017865

Yumura, H., Ashibe, Y., Ohya, M., Itoh, H., Watanabe, M., Masuda, T., Ichikawa, H., Mimura, T., Honjo, S., Hara, T., Ohno, R., Shimoda, M., Nakamura, N., Komagome, T., & Yaguchi, H. (2013). Update of YOKOHAMA HTS cable project. *IEEE Transactions on Applied Superconductivity*, *23*(3), 5402306–5402306. https://doi.org/10.1109/TASC.2013.2245931

Zarbakhsh, N., Misaghian, M. S., & Mcardle, G. (2022). Human mobility-based features to analyse the impact of COVID-19 on power system operation of Ireland. *IEEE Open Access Journal of Power and Energy*, *9*, 213–225. https://doi.org/10.1109/OAJPE.2022.3155960

Zhang, Z., Wang, F., Tolbert, L. M., Blalock, B. J., & Costinett, D. J. (2015). Evaluation of switching performance of SiC devices in PWM inverter-fed induction motor drives. *IEEE Transactions on Power Electronics*, *30*(10), 5701–5711. https://doi.org/10.1109/TPEL.2014.2375827

Zhang, Z.-P., Zheng, C.-J., Zheng, M., Zhao, H., Zhao, J.-K., Sun, W.-F., & Chen, J.-Q. (2020). Interface damages of electrical insulation in factory joints of high voltage submarine cables. *Energies*, *13*(15), 3892. https://doi.org/10.3390/en13153892

Zhong, H., Tan, Z., He, Y., Xie, L., & Kang, C. (2020). Implications of COVID-19 for the electricity industry: A comprehensive review. *CSEE Journal of Power and Energy Systems*, *6*(3), 489–495. https://doi.org/10.17775/CSEEJPES.2020.02500

Zhou, C., Yi, H., & Dong, X. (2017). Review of recent research towards power cable life cycle management. *High Voltage*, *2*(3), 179–187. https://doi.org/10.1049/hve.2017.0037

Pertanika J. Sci. & Technol. 33 (6): 2441 - 2464 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Review Article

A Comprehensive Review of Tuberculosis Detection and Prediction Using Hybrid Machine Learning Models on Ziehl-Neelsen-Stained Microscopy Images

Nur Amirah Kamaluddin[1], Murizah Kassim[1,2]*, Nor Salmah Bakar[3] and Shuria Saaidin[1]

[1]*Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*
[2]*Institute for Big Data Analytics and Artificial Intelligence, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*
[3]*Faculty of Medicine, Universiti Teknologi MARA, 47000 Sungai Buloh, Selangor, Malaysia*

ABSTRACT

Tuberculosis (TB) remains a critical global health challenge, particularly in resource-constrained settings where timely and accurate diagnosis is essential for effective disease management and control. Traditional diagnostic methods, such as Ziehl-Neelsen (ZN)-stained sputum microscopy, are widely employed for detecting *Mycobacterium tuberculosis*; however, these techniques are inherently subjective and prone to variability due to their reliance on manual interpretation. In response, an increasing body of research has applied deep learning (DL)-based approaches to automate TB detection from microscopy images. This systematic review synthesizes findings from 67 studies that have explored various machine-learning techniques for TB diagnosis using ZN-stained images. A structured literature search was conducted across multiple scientific databases, including PubMed, IEEE Xplore, Scopus, and ScienceDirect. Studies were selected based on their focus on DL applications for TB detection using ZN-stained images. The reviewed methodologies encompass various stages, including image preprocessing, feature extraction, classification strategies, and performance evaluation metrics. Our review reveals that DL models, particularly those employing automated feature extraction and classification, are predominantly used, with some studies reporting accuracies of up to 100%. This review provides a comprehensive overview of state-of-the-art methodologies, including image preprocessing, feature extraction, classification strategies, and performance evaluation metrics. Notably, the evidence indicates that convolutional neural network (CNN)-based

ARTICLE INFO

Article history:
Received: 12 February 2025
Accepted: 18 May 2025
Published: 25 September 2025

DOI: https://doi.org/10.47836/pjst.33.6.03

E-mail addresses:
nuramirahkamaluddin97@gmail.com (Nur Amirah Kamaluddin)
murizah@uitm.edu.my (Murizah Kassim)
bsalmah@uitm.edu.my (Nor Salmah Bakar)
shuria6809@uitm.edu.my (Shuria Saaidin)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

approaches offer the highest promise due to their robust ability to detect subtle features in stained images. Consequently, future research will focus on developing and optimizing CNN-based models to further enhance TB detection, ultimately improving diagnostic outcomes and supporting more effective TB control strategies.

Keywords: Convolutional neural network, hybrid machine learning, medical image processing, tuberculosis detection, Ziehl-Neelsen staining

INTRODUCTION

TB remains a significant global health challenge, caused by *Mycobacterium tuberculosis*, which usually affects the lungs but can spread to other organs. According to the World Health Organization (WHO), in 2024, Malaysia reported an incidence of 74 new and relapse TB cases per 100,000 population and 3,140 TB-related deaths. These figures underscore the significant public health challenge posed by TB in the country (WHO, 2024). TB is an airborne infectious illness transmitted when people with active pulmonary TB cough, sneeze, talk, sing, or laugh. It mostly affects the lungs, but it may also impact the spine, brain, and kidneys. Although TB transmission requires prolonged and close contact, its latent form (latent tuberculosis infection or LTBI) allows the bacteria to persist within the host for years without causing symptoms, only becoming active when the immune system is compromised (Tobin & Tristram, 2024). However, comprehensive data on LTBI prevalence in Malaysia remains limited. A study by MacLean et al. (2020) using the tuberculin skin test (TST) reported an LTBI prevalence of 68.20% among inmates in Malaysian prisons, indicating significant latent transmission reservoirs. The high burden of TB and LTBI necessitates advancements in diagnostic approaches to ensure timely detection and treatment, particularly in resource-limited settings where conventional methods face significant challenges. Many individuals with active TB may not immediately realize they are infected, as the disease can sometimes develop slowly over weeks or even months without showing noticeable symptoms (Suliman et al., 2019). When symptoms do appear, they often resemble those of common illnesses such as the flu and may include fever, weight loss, persistent cough, and fatigue. Since TB can take up to six weeks to manifest, delayed detection contributes to continued transmission (Nor et al., 2021). Figure 1 illustrates that an individual with active TB can unknowingly transmit the bacteria to a healthy person, underscoring the urgency for improved early detection methods (Cambier et al., 2014).

TB diagnosis traditionally involves multiple laboratory techniques, including chest radiography, computed tomography (CT) scans, and microbiological assessments of sputum samples. Among these, Ziehl-Neelsen ZN staining for sputum smear microscopy remains one of the most widely used diagnostic techniques in resource-limited settings, as it enables direct visualization of acid-fast bacilli (AFB) under a microscope (Ghosh et al., 2022; Surani et al., 2021). Figure 2 displays an image of TB detected using ZN-stained microscopy.

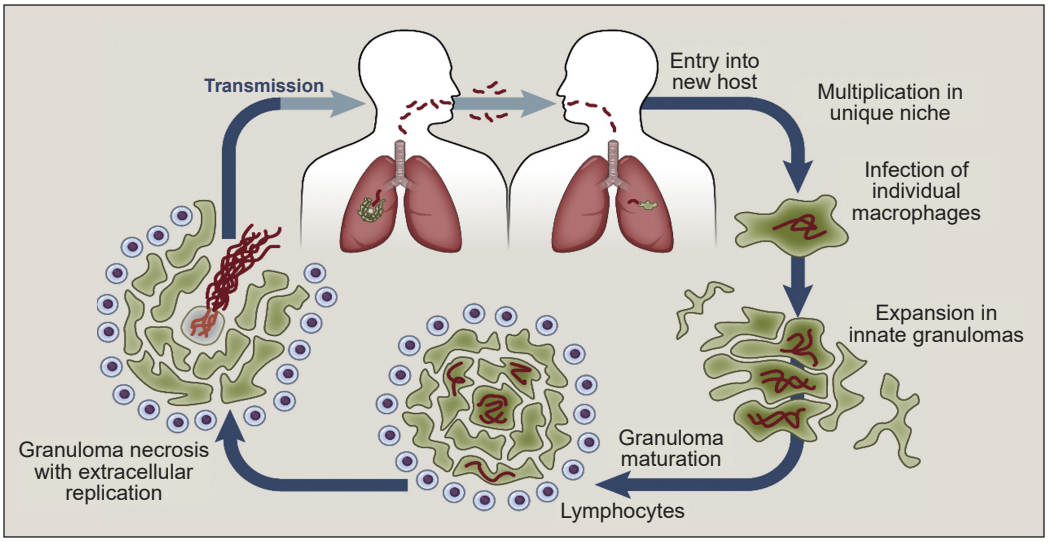


Figure 1. Development of tuberculosis illustration (Cambier et al., 2014)

The image shown below is also displayed. ZN staining is a special bacteriological stain used to identify AFB, which retains the red carbol fuchsin stain even after being washed with acid alcohol. This image is a sample from the Faculty of Medicine laboratory for TB analysis at Universiti Teknologi MARA, Shah Alam. The characteristic appearance of TB bacilli under ZN staining includes curved, rod-shaped bacteria, often wrapped together in cord-like formations. While this staining method provides a direct way to detect TB, it is highly time-consuming since researchers or pathologists must manually examine microscopic fields to locate the tiny bacilli. Given that TB bacteria measure approximately 1-2 micrometers, detecting them requires magnification of at least 40×, making the process labor-intensive and prone to human error. These challenges highlight the urgent need for advanced technologies, like machine learning (ML), to enhance TB diagnosis efficiency and reduce diagnostic errors associated with manual interpretation.

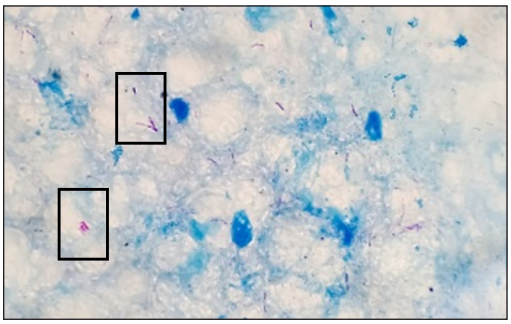


Figure 2. Tuberculosis was detected using Ziehl-Neelsen-stained microscopy under 40× magnification

The application of artificial intelligence (AI) and ML in TB detection has shown considerable potential in improving accuracy, efficiency, and automation (Hooda et al., 2017; Lakhani & Sundaram, 2017). CNNs, a class of DL models specialized in image processing, have demonstrated remarkable performance in medical image classification. Studies have validated CNN-based approaches for TB detection using chest X-rays,

achieving high diagnostic accuracy and significantly reducing the time required for analysis (Mujeeb Rahman et al., 2025; Nafisah & Muhammad, 2024). Sarawagi et al. (2024) further reinforced these findings, demonstrating the potential of CNN architectures in streamlining TB detection workflows in clinical settings. However, while CNN models have been widely applied to radiographic images, their application to ZN-stained sputum smear microscopy images remains underexplored. ZN-stained images present unique challenges such as non-uniform staining, overlapping bacilli, and background noise, necessitating specialized ML techniques tailored for effective feature extraction and classification.

Traditional TB diagnostic methods, such as ZN-stained microscopy, are prone to human error, time-consuming, and suffer from variability in technician expertise, which can impact the accuracy and reliability of diagnoses. While CNNs have been extensively applied to various medical imaging tasks, their specific application to TB detection using ZN-stained tissue sample microscopy images remains underexplored. Moreover, most ML approaches in TB detection lack emphasis on optimizing pre-trained CNN models, such as VGG16, to effectively address the unique challenges associated with ZN-stained microscopy images, including variations in stain intensity and noise artifacts. This review aims to bridge this gap by systematically analyzing and evaluating ML techniques tailored for ZN-stained TB microscopy images, with a focus on pre-trained CNN models and their optimization strategies.

To address these challenges, ML offers a transformative approach for TB diagnosis by automating and standardizing image-based detection. DL models, particularly CNNs, can analyze microscopy images with high precision, reducing the subjectivity associated with manual analysis. Automated image analysis significantly reduces diagnosis time, enabling rapid screening and early intervention. Additionally, ML models eliminate intra- and inter-observer variability, ensuring consistent and reproducible diagnostic outcomes. AI-driven diagnostic frameworks can be integrated into digital pathology systems, enabling remote diagnostics and supporting telemedicine initiatives in resource-constrained regions. Furthermore, this review provides a comparative analysis of existing ML-based approaches for TB detection in ZN-stained images, highlights their strengths and weaknesses, and discusses potential improvements in model training, feature selection, and generalization performance. By systematically evaluating existing ML approaches and their applicability to microscopy-based TB diagnosis, this study aims to bridge the gap between traditional diagnostic techniques and modern computational advancements, ultimately contributing to improved TB detection and disease management.

Traditional TB Diagnosis Approaches

TB remains a daunting worldwide health concern, despite substantial breakthroughs in diagnostic technologies. From a clinical perspective, TB illness is evaluated using a

complete medical assessment that includes a detailed medical history, physical examination, and several types of diagnostic testing (Centers for Disease Control and Prevention [CDC], 2025). These tests typically include tuberculin skin tests (TST) or TB blood tests (interferon-gamma release assays), chest radiography, and laboratory evaluations such as sputum smear microscopy, mycobacterial culture, and molecular tests for drug resistance (Pai et al., 2016). Nevertheless, a definitive diagnosis is based on laboratory procedures that directly detect the presence of tuberculosis germs, most notably sputum smear microscopy and culture. While mycobacterial culture is regarded as the gold standard because of its great sensitivity, it needs an incubation time of up to eight weeks, delaying treatment action (Liang et al., 2022; McClean et al., 2024). Polymerase chain reaction (PCR)-based technologies, such as the GeneXpert MTB/RIF test, provide quick detection and simultaneous drug resistance profiling, but they are restricted by cost and resource restrictions in many endemic areas (Horne et al., 2019).

ZN staining remains an important component of this diagnostic system, especially in resource-limited situations. ZN-stained sputum smear microscopy is commonly used to identify acid-fast bacilli (AFB) associated with *Mycobacterium* TB (Bhandari R, 2021; Masali et al., 2021). The ZN staining method involves several key steps. First, carbol fuchsin is applied to the sputum smear, which penetrates the lipid-rich cell walls of the *Mycobacterium tuberculosis* (TB) bacteria, staining them red. Following this, an acid-alcohol solution is used to decolorize the smear. This step removes the stain from non-acid-fast organisms, leaving only the TB bacteria-stained red. Finally, methylene blue is applied as a counterstain, providing a blue background that contrasts with the red-stained TB bacteria, enhancing their visibility. This causes TB bacteria to appear as brilliant red rods under a light microscope (Bayot et al., 2023; Dzodanu et al., 2019). Figure 3 illustrates the ZN-stained process (LaboratoryInfo, 2022), which is a successful approach that is strongly reliant on the knowledge of the microscopist, resulting in possible inter-observer variability and reduced sensitivity in cases with low bacterial load (Behr et al., 2022; Zaporojan et al., 2024).

Manual examination of ZN-stained smears has been the cornerstone of TB diagnosis in many regions due to its cost-effectiveness and simplicity. However, its accuracy is often compromised by subjective interpretation and inter-observer variability (Perez-Siguas et al., 2023). Despite these limitations, ZN-stained smear microscopy remains critical in high-burden settings that lack advanced diagnostic technologies (Tummalapalli et al., 2024). Automated image analysis systems have been developed to overcome these challenges. These systems employ ML and computer vision techniques to analyze digital images of ZN-stained smears, reducing diagnostic variability, expediting the screening process, and maintaining high sensitivity and specificity (Bhaskar et al., 2023). Additionally, automation facilitates digital archiving and remote consultation, which are valuable for large-scale

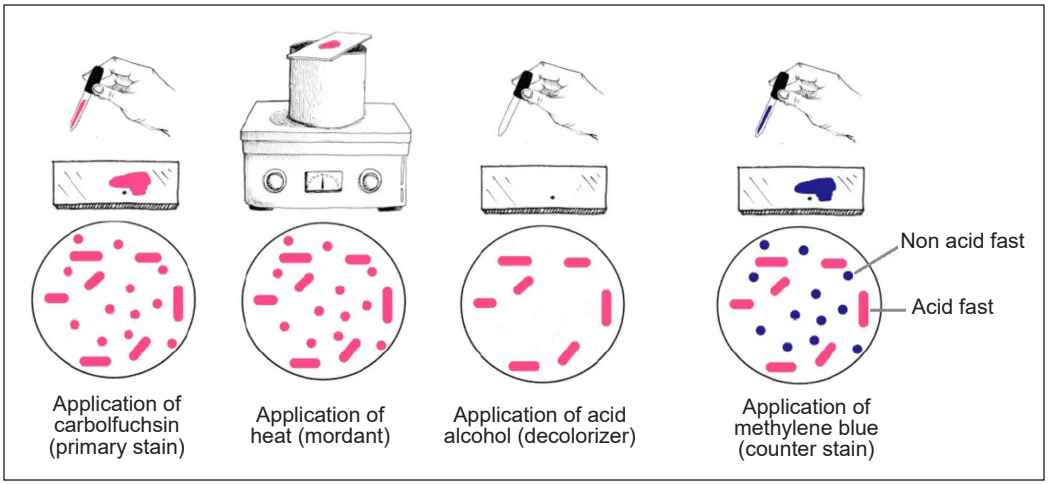


Figure 3. Ziehl-Neelsen-stained process (LaboratoryInfo, 2022)

TB control programs (Shwetha et al., 2021). Although mycobacterial culture and PCR-based assays like ARIMA offer higher sensitivity and rapid drug resistance detection, their high costs and infrastructure requirements limit their widespread use in many endemic areas (Campelo et al., 2021; Li et al., 2022). Therefore, ZN-stained smear microscopy, particularly when enhanced with automated analysis, remains the frontline diagnostic tool for TB screening due to its affordability and rapid turnaround. While previous reviews have broadly examined ML applications in TB detection, most have focused on radiographic images such as chest X-rays or molecular diagnostic methods. However, the application of ML to ZN-stained sputum smear microscopy images remains underexplored. Thus, future research should focus on further integrating automated image analysis with conventional diagnostic workflows to optimize TB control strategies globally.

ML for TB Detection

TB remains a persistent global health challenge, prompting the exploration of innovative diagnostic strategies that extend beyond conventional methods. Traditional techniques such as sputum smear microscopy, mycobacterial culture, and PCR-based assays have been instrumental in TB diagnosis. However, these methods often suffer from limitations. Manual smear microscopy is hindered by subjectivity and inter-observer variability, culture methods are time-consuming, and molecular techniques, though rapid, are expensive and require sophisticated infrastructure (Afsar et al., 2018; Iqbal et al., 2023). In recent years, ML has emerged as a promising tool for improving TB diagnosis. DL techniques, particularly CNN, have revolutionized medical image analysis by automatically learning hierarchical features from complex data. Early research demonstrated the potential of CNN in detecting TB from chest radiographs, with diagnostic accuracies exceeding 90%

(Rajaraman et al., 2022). These encouraging results have spurred further investigation into the use of ML to analyze both radiographic images and ZN-stained sputum smears. However, while most prior reviews primarily focus on ML applications in chest X-ray analysis, limited studies have systematically explored its use in ZN-stained microscopy images, a gap that this review aims to address.

Subsequent studies have expanded on these findings, developed efficient deep network architectures for ZN-stained microscopy images, reporting sensitivities and specificities in the range of 90–95% (Tamura et al., 2024; Witarto et al., 2024). Saini et al. (2023) further refined the approach by integrating segmentation and visualization techniques into CNN frameworks, achieving robust detection accuracies near 93% across varied datasets. These studies illustrate that automated analysis not only improves diagnostic speed but also minimizes human error by providing consistent, objective assessments. Moreover, hybrid approaches combining CNN-based feature extraction with traditional classifiers, such as support vector machines (SVMs) or random forests, have been explored to optimize decision boundaries and further enhance diagnostic performance (Hansun et al., 2023). These methods have shown promise, with several studies reporting accuracies comparable to or even surpassing standalone DL models. Despite these advances, challenges remain. The availability of large, well-annotated datasets is critical for training DL models effectively, and generalizability across diverse populations and imaging conditions is still under investigation (Ahmed et al., 2023). Furthermore, variations in ZN-stained smear characteristics, including stain intensity and background noise, remain a significant hurdle that previous reviews have not extensively analyzed. Addressing these challenges, this study provides a comparative analysis of different CNN architectures, particularly pretrained models like VGG16, and evaluates their suitability for TB detection in ZN-stained microscopy images. Additionally, the review explores optimization strategies for feature extraction and classification, bridging the gap between traditional and AI-driven diagnostic approaches.

Nevertheless, the integration of ML into TB diagnostics represents a significant step forward, with potential applications ranging from automated screening in resource-limited settings to digital archiving and remote consultation for large-scale TB control programs. In summary, the application of machine learning, especially CNNs, has the potential to revolutionize TB detection by offering rapid, accurate, and consistent diagnoses. As ongoing research continues to refine these models and address existing challenges, ML-based approaches are poised to become an integral component of global TB control strategies. This review aims to provide a more targeted analysis of ML applications for ZN-stained images, offering insights into dataset standardization, model generalization, and the real-world feasibility of AI-assisted diagnostics, thereby setting it apart from earlier studies that have focused mainly on radiographic imaging.

Innovations in Medical Diagnostics in Malaysia

The exploration of DL applications in Malaysia has expanded across various fields, particularly in medical diagnostics, agriculture, and engineering, highlighting the growing impact of AI on improving operational efficiency and accuracy. Studies such as Hosain et al. (2024) emphasize the parallels between structured learning approaches and DL model development, particularly in enhancing diagnostic outcomes within the Malaysian healthcare system. Awang et al. (2019) delve into the clinical determinants of severe pulmonary tuberculosis, reinforcing the need for advanced techniques such as CNNs to automate chest radiograph interpretation and improve TB diagnosis. Similarly, Toba et al. (2020) demonstrate that DL models can outperform experienced clinicians in diagnosing congenital heart disease from radiographic images, highlighting the potential of these technologies in improving diagnostic accuracy across various medical conditions in Malaysia. However, most DL studies in Malaysia focus on radiographic imaging (X-rays, CT scans, and magnetic resonance imaging [MRI]), such as Kotei and Thirunavukarasu (2024), while few have addressed the complexity of ZN-stained sputum smear microscopy. Given the unique challenges posed by these images, such as variations in staining intensity and the presence of artifacts, specialized CNN architectures are required to optimize detection performance.

Beyond the healthcare sector, Lu et al. (2020) explore the efficacy of DL methods like LSTMs combined with fully convolutional networks (FCNs) in brain signal processing, underscoring their applicability to other domains, such as neurotechnology and healthcare in Malaysia. Carvalho et al. (2023) further contribute to this narrative by demonstrating how DL models can enhance feature classification accuracy, which could be applied to school health services, ensuring better health outcomes for children in Malaysia. However, unlike large-scale chest radiograph datasets, annotated ZN-stained smear microscopy datasets are scarce in Malaysia. This necessitates innovative approaches such as transfer learning, data augmentation, or synthetic dataset generation to improve model robustness and generalizability across diverse clinical settings.

In medical imaging, Alaskar et al. (2019) show how DL models, such as Alex Net and Google Net, can effectively detect ulcers in wireless capsule endoscopy (WCE) images, offering valuable insights for broader healthcare applications, including TB diagnostics. Xie et al. (2020) present a novel approach using a fully CNNs to detect pulmonary tuberculosis lesions, achieving impressive diagnostic metrics, which are crucial for improving early TB detection in Malaysia. The use of ensemble DL models for TB detection, as discussed by Hwa et al. (2019). It also highlights the importance of leveraging innovative semi-automated methods to differentiate between clinically pulmonary TB and lung cancer, thus contributing to the effective management of TB in high-prevalence regions. However, while CNNs have demonstrated effectiveness in various Malaysian healthcare applications,

their direct applicability to TB detection in ZN-stained images remains underexplored. Optimizing pre-trained models such as VGG16 specifically for smear microscopy is crucial to improving sensitivity and specificity in this context.

Finally, Tiwari et al. (2023) extend the application of DL techniques into engineering, specifically in fault diagnosis, emphasizing model robustness through dropout techniques, which is essential for improving diagnostic accuracy across diverse sectors in Malaysia. Collectively, these studies highlight the transformative potential of DL in Malaysia, particularly in enhancing diagnostic accuracy, operational efficiency, and disease management in healthcare, while also contributing to advancements in other critical fields, such as agriculture and engineering. Moreover, AI adoption in Malaysia is still evolving, with regulatory and infrastructural challenges. Discussing how automated TB detection could be integrated into Malaysia's public health policies or screening programs (e.g., leveraging mobile diagnostics for rural areas) enhances the practical relevance of AI-driven approaches. These findings highlight the role of DL technologies in addressing the unique challenges faced by Malaysia, positioning AI as a key tool in enhancing diagnostic performance and public health outcomes.

METHODOLOGY

The analysis includes studies on tuberculosis and machine learning, with a focus on "ZN-stained". The queries returned to several journal and conference publications. This survey solely considers peer-reviewed literature for systematic reviews. Table 1 shows the search engines used, which included Google Scholar, Elsevier, and Springer. The number of search results indicates the number of articles retrieved by search engines for the specified keywords. The number of relevant articles denotes the number of items that passed the initial screening procedure and were judged possibly relevant for a thorough assessment. The bibliographic part of the papers was also examined. The method was iterated until no further items were located.

Prospective research articles are identified, screened, and chosen based on their eligibility. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) model, depicted in Figure 4, outlines the overall publications reviewed in the research. It shows a step-by-step flow chart for the detection technique. Papers that were unrelated to ZN-stained (100 articles) were not considered. After examining 156 publications, 39 were

Table 1
Summary of search results and retrieved relevant articles

Database engines	Number of searches	Number of relevant articles
Elsevier	1,006	20
Springer	217	10
Pubmed	240	26
IEEE Xplore	100	6
MDPI	200	4t
Taylor & Francis	800	10

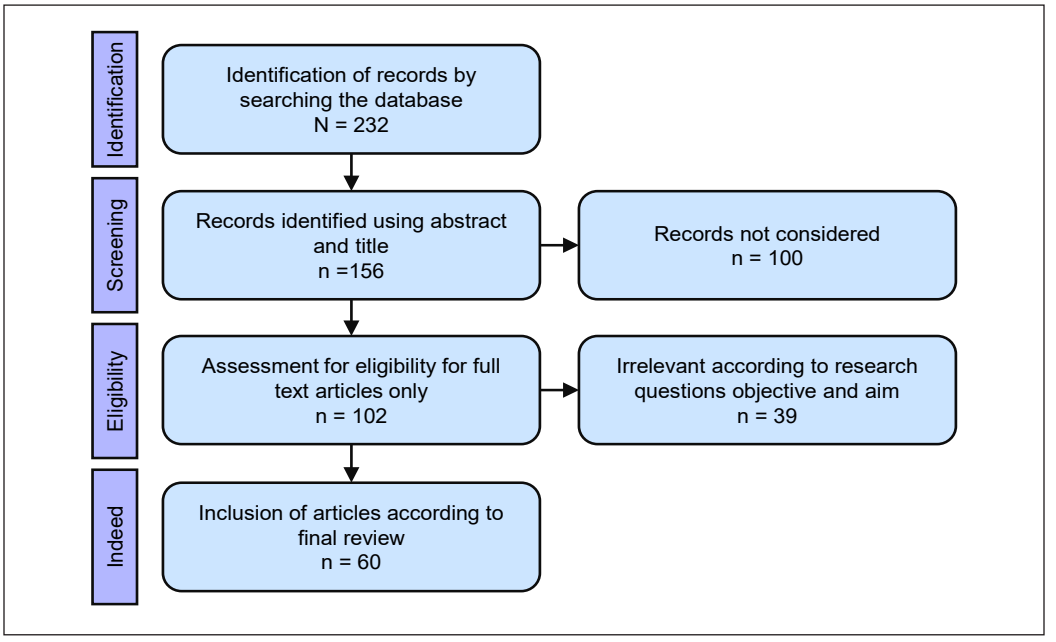


Figure 4. PRISMA model for the depiction of inclusion and exclusion of records

discarded based on title and abstract analysis. The preprint version and duplicate publications were also eliminated. After assessing the quality of published research, 102 papers were selected, with 150 discarded for not being research articles or relevant concepts. As a result, 60 papers were chosen for a detailed assessment.

Tuberculosis Detection and Prediction Framework

A detailed methodology for tuberculosis detection and prediction using a hybrid ML approach is presented in a framework. Figure 5 shows that the framework is structured into four essential phases. In Phase 1, data collection and initial processing occur, during which images of acid-fast bacilli are obtained from the clinical laboratory at Hospital Puncak Alam. Phase 2, referred to as data development, involves refining and transforming the raw dataset. This stage focuses on constructing a well-organized dataset suitable for both training and validating predictive models by applying preprocessing techniques, such as normalization, feature engineering, and data augmentation, to optimize model performance. Phase 3 represents the core of the framework, where the prediction model is trained and subsequently validated. Here, a CNN utilizing the VGG16 architecture is employed. CNNs are particularly adept for this task due to their ability to discern subtle color differences in images, which is a crucial factor, given that the ZN staining method exploits the unique properties of the bacterial cell wall, rich in mycolic acid, rendering it resistant to decolorization by acid-alcohol. This staining technique is critical for diagnosing

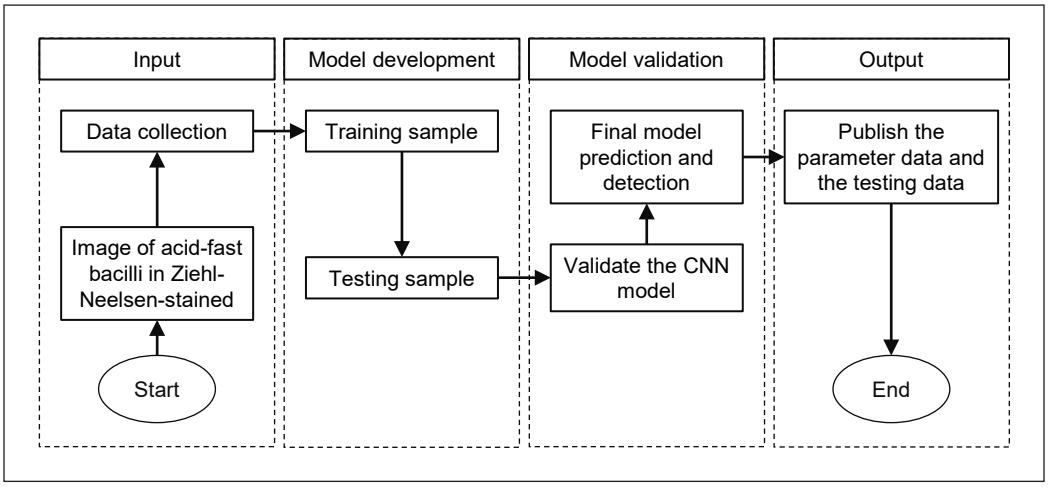


Figure 5. The framework for modelling the tuberculosis detection and prediction on Ziehl-Neelsen-stained slides using hybrid machine learning
Note. CNN = Convolutional neural network; TB = Tuberculosis; ZN = Ziehl-Neelsen

TB in clinical samples like sputum smears or tissue sections. Finally, the CNN model leverages the meticulously curated and preprocessed dataset from Phase 2 to perform accurate TB detection and prediction.

RESULTS AND DISCUSSION

Data Collection

The data for this study were obtained from the Clinical Diagnostic Laboratories (CDL) at Hospital UiTM Puncak Alam. ZN staining is a widely used technique in microbiology and pathology for detecting acid-fast bacilli, particularly *Mycobacterium tuberculosis*. This method capitalizes on the high mycolic acid content in bacterial cell walls, rendering them resistant to acid-alcohol decolorization.

Despite its diagnostic importance, analyzing ZN-stained samples remains a labor-intensive process, requiring skilled professionals for accurate interpretation. Figure 6 presents a sample of the dataset used in this study. However, one of the main challenges in working with ZN-stained images is their inherent variability due to differences in staining intensity, image contrast, and sample preparation techniques. Addressing these inconsistencies is crucial

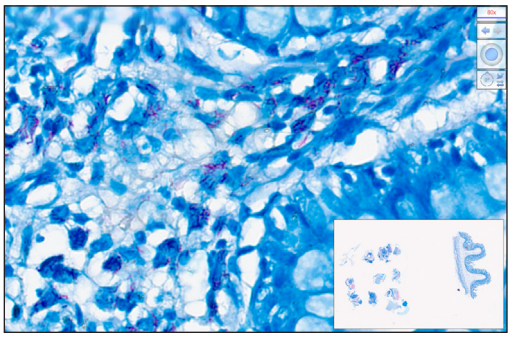


Figure 6. One of the Ziehl-Neelsen-stained microscopy images under 80× magnification using NDP.view2 software

for improving the robustness of automated TB detection models. Since the bacilli are extremely small, measuring approximately 1 μm in size, a magnification of 40× or higher was required to visualize them. As a result, a single glass slide could contain thousands of images, with the bacilli appearing as red-colored rods. However, capturing and saving these images was a time-consuming process, requiring meticulous effort to ensure proper documentation and storage.

Comparison of Detecting TB Using ML

Recent advancements in ML have paved the way for automating TB detection, significantly reducing the burden on healthcare professionals. CNNs have demonstrated exceptional performance in medical image analysis by autonomously learning hierarchical features from complex datasets. Initial studies on chest radiographs reported diagnostic accuracies exceeding 90%, motivating further exploration of ML techniques for TB detection (Lakhani & Sundaram, 2017).

Table 2 summarizes various ML models applied to TB detection. While many studies focus on chest X-rays due to their rapid image acquisition and accessibility, relatively few have explored the application of ML on ZN-stained images. The comparative analysis reveals that most studies have leveraged CNN-based models, such as VGG16, ResNet, and EfficientNet, achieving high accuracy levels. However, a key limitation is that ZN-stained sample analysis requires extensive manual effort, limiting its widespread adoption in automated diagnostic workflows. Despite this, automating ZN-stained smear analysis could provide substantial benefits, particularly in resource-limited settings where access to advanced imaging modalities is constrained.

Among the different ML architectures applied to TB detection, CNN-based models, particularly VGG16, have consistently demonstrated superior performance. The high accuracy of CNNs can be attributed to their ability to automatically extract and learn hierarchical features from image data, reducing reliance on manual feature engineering. However, deeper models such as ResNet and EfficientNet, while offering better feature extraction, tend to require more computational power, making them less suitable for real-time clinical deployment in resource-limited settings. Additionally, studies reporting exceptionally high accuracy (up to 100%) raise concerns about potential overfitting, particularly if models have been trained on small datasets without proper validation. Future work should ensure rigorous evaluation using large-scale, diverse datasets and external validation cohorts to confirm the generalizability of these models.

Challenges in ZN-Stained Image Processing

ZN-stained images present unique challenges for automated analysis compared to chest X-rays. Variability in staining intensity, contrast, and background noise can impact model

Table 2
Automated image detection

Method	Sample	Accuracy (%)	Novelty
ResNet and EfficientNet (Munadi et al., 2020)	Chest X-ray	89.92 and 94.80	Evaluating the impact of three image enhancement techniques (UM, HEF, and CLAHE) on improving the performance of DL models (ResNet and EfficientNet) for tuberculosis detection in chest X-ray images
CNNs (DenseNet201) (Rahman et al., 2020)	Chest X-ray	96.47	Demonstrating that tuberculosis detection accuracy improves when using segmented lung images rather than whole chest X-ray images, with DenseNet201 achieving superior performance
SVM (Hrizi et al., 2022)	Chest X-ray	~93	Optimizing tuberculosis detection by integrating a genetic algorithm (GA) for feature selection with a modified SVM classifier
Image segmentation (Al-Timemy et al., 2021)	Chest X-ray – TB, Covid-19	91.60	Developing a computationally efficient pipeline using ResNet-50 for deep feature extraction and an ensemble of subspace discriminators to accurately distinguish COVID-19, TB, and other chest infections from X-ray images
CNN (Guo et al., 2020)	Chest X-ray	98.46	Integrating CNN modifications, artificial bee colony algorithm-based fine-tuning, and a linear average-based ensemble method to enhance TB detection and localization in CXRs while also differentiating among seven TB-related manifestations using class activation mapping for visual interpretability
Automated image analysis using CNN and traditional ML (Huang et al., 2022)	ZN-stained smear	88–90	Evaluating the μ-Scan 2.0 automation system in a multi-center, double-blind trial, demonstrating its potential to enhance TB smear microscopy sensitivity, efficiency, and remote accessibility for resource-limited areas
CNN (Panicker et al., 2018)	ZN-stained smear	97.13	An automatic TB detection method from sputum smear images using image binarization and CNN-based classification, improving diagnostic speed and accuracy compared to conventional microscopy
CNN-based model (Tasci et al., 2021)	Chest X-ray	97.50	Proposing a voting and preprocessing variations-based ensemble CNN model incorporating Bayesian optimization-weighted soft voting and CLAHE-enhanced image transformations, achieving state-of-the-art accuracy in TB detection from CXR images.
CNN (Rajaraman & Antani, 2020)	Chest X-ray	95	Leveraging modality-specific knowledge transfer from large-scale chest X-ray datasets and integrating a stacked ensemble of top-performing DL models to enhance TB detection accuracy
U-Net segmentation (Rajaraman et al., 2021)	Chest X-ray	~90	Training CXR modality-specific U-Net models for semantic segmentation of TB-consistent findings and enhancing segmentation performance by augmenting training data with weak TB-consistent localizations derived from a DL classifier

Note. CNN = Convolutional neural network; TB = Tuberculosis; ZN = Ziehl-Neelsen; DL = Deep learning; UM = Unsharp masking; HEF = Histogram equalization filtering; CLAHE = Contrast Limited Adaptive Histogram Equalization; SVM = Support vector machine; GA = Genetic algorithm; CXR = Chest X-ray; AI = Artificial intelligence; U-Net = U-shaped convolutional network

performance, leading to false positives or negatives (Shah et al., 2017). Traditional image preprocessing techniques such as contrast enhancement, histogram equalization, and denoising can mitigate these challenges to some extent. However, DL-based preprocessing methods, such as generative adversarial networks (GANs) for image enhancement or self-supervised learning for feature extraction, could provide more robust solutions. Another major limitation is the presence of overlapping bacilli in ZN-stained smears, which can confuse segmentation algorithms. Advanced segmentation models, such as U-Net and Mask R-CNN, can be integrated into ML pipelines to enhance bacillus localization and classification accuracy.

Clinical Integration and Real-World Applications

Although CNN-based models have demonstrated high accuracy, their integration into clinical workflows remains a challenge. One major barrier is the lack of model interpretability; clinicians need to understand why a model made a certain prediction. Explainability techniques such as gradient-weighted class activation mapping (Grad-CAM) and SHapley Additive exPlanations (SHAP) could enhance trust in AI-assisted TB diagnosis by visualizing which image regions contributed to the model's decision (Narkhede, 2024). Additionally, regulatory considerations and ethical concerns must be addressed before deploying AI-based TB detection in real-world settings. Ensuring fairness and avoiding biases in AI models is crucial, especially when datasets are skewed towards specific populations or biased by certain staining methods. Implementing human-in-the-loop AI systems, where models assist rather than replace clinicians, could strike a balance between automation and medical expertise.

While CNNs have dominated TB detection research, alternative approaches could further enhance performance. Hybrid models that combine CNNs with traditional ML classifiers, such as SVMs or random forests, have shown promise in improving diagnostic accuracy (Narkhede, 2024). Moreover, transformer-based architectures, such as vision transformers (ViTs), could offer superior feature representation for medical image analysis, though their effectiveness on ZN-stained images remains largely unexplored. Self-supervised learning, where models learn representations from unlabeled data, could also be beneficial, especially given the limited availability of annotated ZN-stained datasets. This approach has the potential to reduce dependency on large, labeled datasets while improving model robustness.

Review of TB Detection on ZN-Stained Using ML

The review of TB detection using ZN-stained microscopy images through ML is comprehensively summarized in Table 3. This review specifically focuses on research studies that have utilized ZN-stained smear images as the primary data source for TB

detection, in contrast to the more extensively studied chest X-ray (CXR) images. While CXRs have been widely adopted in TB screening due to their accessibility and ease of acquisition, ZN-stained smear microscopy remains the gold standard for bacteriological confirmation of TB, particularly in low-resource settings. However, manual examination of these smears by trained microbiologists is time-consuming, labor-intensive, and prone to interobserver variability, leading to potential inconsistencies in diagnosis.

The table provides an overview of key research works in this domain, detailing essential aspects such as the author(s), the type of sample used (i.e., ZN-stained sputum smear images), the ML techniques and architectures employed, including CNNs, U-Net, hybrid DL models, and other automated classification methods—and the reported performance metrics, such as accuracy, sensitivity, and specificity. One of the primary challenges associated with analyzing ZN-stained microscopy images stems from their inherent characteristics, including low contrast, background noise, staining inconsistencies, and the presence of artifacts, all of which can hinder the performance of traditional image-processing approaches. Despite these challenges, the reviewed studies demonstrate that well-designed machine-learning pipelines can achieve promising diagnostic accuracy, typically within the range of 88 to 90%.

These findings suggest that automated TB detection from ZN-stained smears has significant potential for clinical applications, particularly as a supportive tool to aid pathologists and laboratory technicians in making diagnostic decisions. By leveraging DL models trained on large-scale annotated datasets, such systems can enhance diagnostic efficiency, reduce human error, and improve consistency across different laboratory settings. Furthermore, the integration of ML-based TB detection into routine clinical workflows could be particularly transformative in resource-limited regions where access to skilled personnel and advanced diagnostic tools remains a significant barrier to timely and accurate TB diagnosis. Table 3 presents a summary of relevant studies, highlighting their methodologies, datasets, and outcomes, thereby providing a comparative analysis of the current state in this field. This review underscores the growing importance of AI in infectious disease diagnostics and reinforces the notion that automated image-based TB detection could serve as a valuable adjunct to existing diagnostic methods, ultimately contributing to global TB control efforts.

In summary, our review indicates that among the various ML models applied to TB detection using ZN-stained images, CNN-based approaches have demonstrated the highest performance. Some studies even report an accuracy of up to 100%. This remarkable level of accuracy not only exceeds that of traditional ML methods and hybrid models but also highlights the robustness of CNN in automatically extracting and processing complex image features. Such outstanding performance provides a strong rationale for focusing our research on CNN-based models, as they offer a reliable and efficient solution for automated TB diagnosis.

Table 3
Tuberculosis detection on Ziehl-Neelsen-stained slides using machine learning

Method / Technique	Samples	Accuracy (%)	Novelty
CNN - VGG16, ResNet50, and SqueezeNet (Shwetha et al., 2021)	200 images	97	The usage of SqueezeNet for bacilli detection, which achieves 97% accuracy with a lightweight model, is more efficient than VGG16 and ResNet50 while maintaining excellent performance
Coarse (RGB) and fine (Sauvola) level segmentation (Samuel & Baskaran, 2021)	Not described	98.70	Integration of RGB thresholding and Sauvola's adaptive thresholding for TB bacilli segmentation, combined with shape descriptors for precise feature extraction, enables an automated and efficient TB detection system that reduces manual effort while improving sensitivity and specificity
Pat-Scan, scanner, and software (Sua et al., 2021)	2,000 images	99	Developing a digital pathology program for detecting and quantifying both typical and atypical mycobacteria in paraffin-embedded ZN-stained tissues
HSV color space transformation and image segmentation (Riza et al., 2022)	51 images	78.68	A freely available, diverse dataset supporting autofocusing, auto stitching, and bacilli segmentation, which enhances algorithm development and validation for automated TB detection
CNN (Zaizen et al., 2022)	Not described	98	Detecting AFB in bronchoscopy samples demonstrates significantly higher sensitivity (86%) compared to conventional bacteriological tests (29%) for the TB diagnosis
CNN - AlexNet, VGGNet-19, ResNet-18, DenseNet, GoogLeNet-incept-v3, In-ceptionResNet-v2, and the classic three-layer model (Shelomentseva & Chentsov, 2021)	Not described	99	A simple three-layer convolutional neural network outperforms advanced transfer learning models like DenseNet and InceptionResNet-v2 for TB bacilli detection in ZN-stained images, highlighting the potential of lightweight CNNs for automated diagnosis
CNN - ResNet-18, ResNet-50, and VGG-16 (Rachmad, 2024)	5,100 images	93.42	Demonstrating that AlexNet outperforms ResNet-18, ResNet-50, and VGG-16 in both accuracy (93.42%) and processing speed (5 min 52 s) for TB detection
CNN (Yang et al., 2020)	21,504 images	87.62	Developing an ML pipeline that combines two CNN models with an active learning framework and logistic regression, achieving high sensitivity (87.13%) and specificity (87.62%) for AFB detection
RegNetX4 (Zurac et al., 2022)	510 images	98.33	Developing a high-performance AI-based mycobacteria identification method using a large dataset of over 260,000 positive and 700 million negative patches from 510 ZN-stained whole slide images

Table 3 *(continue)*

Method / Technique	Samples	Accuracy (%)	Novelty
CNN and SVM (Rachmad et al., 2020)	1,000 images	97.60, 97.90	ResNet-101 architecture combined with SVM for TB bacteria classification in ZN-stained images
CNN – VGG16 (Author's ongoing work)	~1,000 images	-	Developing a hybrid DL model for tuberculosis detection using ZN-stained microscopy images, integrating CNN architectures (such as VGG16) while also exploring preprocessing and segmentation techniques to enhance detection performance

Note. CNN = Convolutional neural network; TB = Tuberculosis; ZN = Ziehl-Neelsen; RGB = Red Green Blue; AFB = Acid-fast bacilli; HSV = Hue Saturation Value; SVM = Support vector machine; DL = Deep learning; ML = Machine learning; Pat-Scan = Pathology scanner; Grad-CAM = Gradient-weighted class activation mapping; SHAP = SHapley Additive exPlanations: RESNET = Residual network; VGG = Visual Geometry Group; DenseNet = Densely connected convolutional network

Despite significant advancements in ML for TB detection, several challenges remain. First, dataset biases, particularly in staining methods and image acquisition settings, can affect model performance across different institutions. Future research should focus on developing standardized datasets that encompass a diverse range of staining variations to improve model generalization. Second, the lack of large-scale, publicly available ZN-stained image datasets hampers progress in this field. Establishing collaborative research initiatives to share anonymized datasets could facilitate better benchmarking and validation of ML models. Federated learning, where models are trained across multiple hospitals without sharing raw data, could be a promising approach to overcome privacy concerns while improving model robustness. Lastly, real-time deployment of ML models in clinical settings requires lightweight architectures that balance accuracy with computational efficiency. Future research should explore model compression techniques such as knowledge distillation or quantization to make CNN-based models more suitable for deployment in resource-limited environments.

CONCLUSION

This comprehensive analysis researched 67 studies relating to machine learning-based TB identification with ZN-stained microscope images. The results show that CNNs are the most successful strategy, consistently outperforming classic ML models in feature extraction and classification. The capacity of CNNs to identify minute patterns in microscope pictures makes them an effective tool for automated tuberculosis diagnosis. Furthermore, this paper emphasizes the importance of hybrid models, which combine CNNs with classical classifiers such as SVMs or VGG16, demonstrating the potential for future performance gains. The study also highlights the importance of image preprocessing methods, such

as contrast enhancement and segmentation, in improving model accuracy. Compared to prior reviews, this paper presents a more targeted investigation of ML algorithms applied to ZN-stained microscopy pictures, shedding light on the strengths and limits of various approaches.

Despite these developments, significant hurdles remain to establishing scalable and clinically effective AI models for tuberculosis diagnosis. Future research should focus on increasing dataset diversity to ensure models generalize across varied imaging conditions and patient populations. Hybrid and explainable AI technologies can help to improve model interpretability and clinician trust. Furthermore, real-world clinical validation and deployment methodologies must be investigated to integrate AI-based TB diagnosis into regular diagnostics. Addressing these problems will be crucial in enhancing the role of ML in TB screening, ultimately leading to more accurate and efficient disease diagnosis in global healthcare settings.

ACKNOWLEDGEMENTS

The authors acknowledge the College of Engineering and the Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Universiti Teknologi MARA, Shah Alam, Selangor, for the support fund provided for the publication of this paper.

REFERENCES

Afsar, I., Gunes, M., Er, H., & Sener, A. G. (2018). Comparison of culture, microscopic smear and molecular methods in diagnosis of tuberculosis. *Revista Española de Quimioterapia*, *31*(5), 435–438.

Ahmed, I. A., Senan, E. M., Shatnawi, H. S. A., Alkhraisha, Z. M., & Al-Azzam, M. M. A. (2023). Multi-techniques for analyzing X-ray images for early detection and differentiation of pneumonia and tuberculosis based on hybrid features. *Diagnostics*, *13*(4), 814. https://doi.org/10.3390/diagnostics13040814

Al-Timemy, A. H., Khushaba, R. N., Mosa, Z. M., & Escudero, J. (2021). An efficient mixture of deep and machine learning models for COVID-19 and tuberculosis detection using X-Ray images in resource limited settings. In D. Oliva, S. A. Hassan, & A. Mohamed (Eds.), *Artificial intelligence for COVID-19* (Vol. 358, pp. 77–100). Springer. https://doi.org/10.1007/978-3-030-69744-0_6

Alaskar, H., Hussain, A., Al-Aseem, N., Liatsis, P., & Al-Jumeily, D. (2019). Application of convolutional neural networks for automated ulcer detection in wireless capsule endoscopy images. *Sensors*, *19*(6), 1265. https://doi.org/doi:10.3390/s19061265

Awang, H., Nik Husain, N., & Abdullah, H. (2019). Chest radiographic findings and clinical determinants for severe pulmonary tuberculosis among children and adolescents in Malaysia. *Russian Open Medical Journal*, *8*(2), e210. https://doi.org/10.15275/rusomj.2019.0210

Bayot, M. L., Mirza, T. M., & Sharma, S. (2023). *Acid fast bacteria*. StatPearls Publishing.

Behr, M. A., Lapierre, S. G., Kunimoto, D. Y., Lee, R. S., Long, R., Sekirov, I., Soualhine, H., & Turenne, C. Y. (2022). Chapter 3: Diagnosis of tuberculosis disease and drug-resistant tuberculosis. *Canadian*

Journal of Respiratory, Critical Care, and Sleep Medicine, *6*(sup1), 33–48. https://doi.org/10.1080/24745332.2022.2035638

Bhandari, R., Gaur, D. S., Kotwal, A., & Kusum, A. (2021). Comparison of Ziehl-Neelsen (ZN) staining and fluorescent (FL) staining in suspected cases of tuberculosis. *International Journal of Pathology and Clinical Research*, *7*(1), 122. https://doi.org/10.23937/2469-5807/1510122

Bhaskar, N., Waghmare, P. T., & Sairam, V. A. (2023). Deep learning framework for automated drug resistance prediction of tuberculosis using computed tomography images. In *International Conference on Next Generation Electronics* (pp. 1–5). IEEE. https://doi.org/10.1109/NEleX59773.2023.10421083

Cambier, C. J., Falkow, S., & Ramakrishnan, L. (2014). Host evasion and exploitation schemes of *Mycobacterium tuberculosis*. *Cell*, *159*(7), 1497–1509. https://doi.org/10.1016/j.cell.2014.11.024

Campelo, T. A., de Sousa, P. R. C., de Lima Nogueira, L., Frota, C. C., & Antas, P. R. Z. (2021). Revisiting the methods for detecting *Mycobacterium tuberculosis*: What has the new millennium brought thus far? *Access Microbiol*, *3*(8), 000245. https://doi.org/10.1099/acmi.0.000245

Carvalho, T. F. M., Santos, V. L. A., Silva, J. C. F., de Assis Figueredo, L. J., de Miranda, S. S., de Oliveira Duarte, R., & Guimarães, F. G. (2023). A systematic review and repeatability study on the use of deep learning for classifying and detecting tuberculosis bacilli in microscopic images. *Progress in Biophysics and Molecular Biology*, *180–181*, 1–18. https://doi.org/10.1016/j.pbiomolbio.2023.03.002

Centers for Disease Control and Prevention. (2025). *Clinical and laboratory diagnosis for tuberculosis*. CDC. https://www.cdc.gov/tb/hcp/testing-diagnosis/clinical-and-laboratory-diagnosis.html

Dzodanu, E. G., Afrifa, J., Acheampong, D. O., & Dadzie, I. (2019). Diagnostic yield of fluorescence and Ziehl-Neelsen staining techniques in the diagnosis of pulmonary tuberculosis: A comparative study in a district health facility. *Tuberculosis Research and Treatment*, *2019*, 4091937. https://doi.org/10.1155/2019/4091937

Ghosh, S., Felix, D., Kammerer, J. S., Talarico, S., Brostrom, R., Starks, A., & Silk, B. (2022). Evaluation of sputum-culture results for tuberculosis patients in the United States-affiliated Pacific Islands. *Asia Pacific Journal of Public Health*, *34*(2-3), 258–261. https://doi.org/10.1177/10105395211060119

Guo, R., Passi, K., & Jain, C. K. (2020). Tuberculosis diagnostics and localization in chest X-rays via deep learning models. *Frontiers in Artificial Intelligence*, *3*, 583427. https://doi.org/10.3389/frai.2020.583427

Hansun, S., Argha, A., Liaw, S.-T., Celler, B. G., & Marks, G. B. (2023). Machine and deep learning for tuberculosis detection on chest X-rays: Systematic literature review. *Journal of Medical Internet Research*, *25*, e43154. https://doi.org/10.2196/43154

Hooda, R., Sofat, S., Kaur, S., Mittal, A., & Meriaudeau, F. (2017). Deep-learning: A potential method for tuberculosis detection using chest radiography. In *IEEE International Conference on Signal and Image Processing Applications* (pp. 497–502). IEEE. https://doi.org/10.1109/ICSIPA.2017.8120663

Horne, D. J., Kohli, M., Zifodya, J. S., Schiller, I., Dendukuri, N., Tollefson, D., Schumacher, S. G., Ochodo, E. A., Pai, M., & Steingart, K. R. (2019). Xpert MTB/RIF and Xpert MTB/RIF Ultra for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database of Systematic Reviews*, (6), CD009593. https://doi.org/10.1002/14651858.CD009593.pub4

Hosain, M. T., Jim, J. R., Mridha, M. F., & Kabir, M. M. (2024). Explainable AI approaches in deep learning: Advancements, applications and challenges. *Computers and Electrical Engineering*, *117*, 109246. https://doi.org/10.1016/j.compeleceng.2024.109246

Hrizi, O., Gasmi, K., Ltaifa, I. B., Alshammari, H., Karamti, H., Krichen, M., Ammar, L. B., & Mahmood, M. A. (2022). Tuberculosis disease diagnosis based on an optimized machine learning model. *Journal of Healthcare Engineering*, *2022*, 8950243. https://doi.org/10.1155/2022/8950243

Huang, H.-C., Kuo, K.-L., Lo, M.-H., Chou, H.-Y., & Lin, Y. E. (2022). Novel TB smear microscopy automation system in detecting acid-fast bacilli for tuberculosis - A multi-center double blind study. *Tuberculosis*, *135*, 102212. https://doi.org/10.1016/j.tube.2022.102212

Hwa, S. K. T., Hijazi, M. H. A., Bade, A., Yaakob, R., & Jeffree, M. S. (2019). Ensemble deep learning for tuberculosis detection using chest X-ray and Canny edge detected images. *IAES International Journal of Artificial Intelligence*, *8*(4), 429–435. https://doi.org/10.11591/ijai.v8.i4.pp429-435

Iqbal, A., Usman, M., & Ahmed, Z. (2023). Tuberculosis chest X-ray detection using CNN-based hybrid segmentation and classification approach. *Biomedical Signal Processing and Control*, *84*, 104667. https://doi.org/10.1016/j.bspc.2023.104667

Kotei, E., & Thirunavukarasu, R. (2024). Tuberculosis detection from chest X-ray image modalities based on transformer and convolutional neural network. *IEEE Access*, *12*, 97417–97427. https://doi.org/10.1109/ACCESS.2024.3428446

LaboratoryInfo. (2022). *Ziehl-Neelsen stain (ZN-stain): Principle, procedure, reporting and modifications*. https://laboratoryinfo.com/zn-stain/

Lakhani, P., & Sundaram, B. (2017). Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, *284*(2), 574–582. https://doi.org/10.1148/radiol.2017162326

Li, J., Ouyang, J., Yuan, J., Li, T., Luo, M., Wang, J., & Chen, Y. (2022). Establishment and evaluation of an overlap extension polymerase chain reaction technique for rapid and efficient detection of drug-resistance in *Mycobacterium tuberculosis*. *Infectious Diseases of Poverty*, *11*, 31. https://doi.org/10.1186/s40249-022-00953-5

Liang, S., Ma, J., Wang, G., Shao, J., Li, J., Deng, H., Wang, C., & Li, W. (2022). The application of artificial intelligence in the diagnosis and drug resistance prediction of pulmonary tuberculosis. *Frontiers in Medicine*, *9*, 935080. https://doi.org/10.3389/fmed.2022.935080

Lu, J., Yan, H., Chang, C., & Wang, N. (2020). Comparison of machine learning and deep learning approaches for decoding brain computer interface: An fNIRS study. In Z. Shi, S. Vadera, & E. Chang (Eds.), *Intelligent Information Processing X: 11th IFIP TC 12 International Conference* (pp. 192–201). Springer. https://doi.org/10.1007/978-3-030-46931-3_18

MacLean, E., Kohli, M., Weber, S. F., Suresh, A., Schumacher, S. G., Denkinger, C. M., & Pai, M. (2020). Advances in molecular diagnosis of tuberculosis. *Journal of Clinical Microbiology*, *58*(10), e01582-19. https://doi.org/doi:10.1128/jcm.01582-19

Masali, H. T., Takpere, A., & Shahapur, P. (2021). A comparative study of Ziehl-Neelsen stain and fluorescent stain microscopy in the diagnosis of pulmonary tuberculosis. *Journal of Pure and Applied Microbiology*, *15*(4), 2027–2033. https://doi.org/10.22207/JPAM.15.4.24

McClean, M., Panciu, T. C., Lange, C., Duarte, R., & Theis, F. (2024). Artificial intelligence in tuberculosis: A new ally in disease control. *Breathe*, *20*(3), 240056. https://doi.org/10.1183/20734735.0056-2024

Mujeeb Rahman, K. K., Zulaikha, S., Dhafer, B., & Ahmed, R. (2025). Advancing tuberculosis screening: A tailored CNN approach for accurate chest X-ray analysis and practical clinical integration. *Intelligence-Based Medicine*, *11*, 100196. https://doi.org/10.1016/j.ibmed.2024.100196

Munadi, K., Muchtar, K., Maulina, N., & Pradhan, B. (2020). Image enhancement for tuberculosis detection using deep learning. *IEEE Access*, *8*, 217897–217907. https://doi.org/10.1109/ACCESS.2020.3041867

Nafisah, S. I., & Muhammad, G. (2024). Tuberculosis detection in chest radiograph using convolutional neural network architecture and explainable artificial intelligence. *Neural Computing and Applications*, *36*, 111–131. https://doi.org/10.1007/s00521-022-07258-6

Narkhede, J. (2024). Comparative evaluation of post-hoc explainability methods in AI: LIME, SHAP, and Grad-CAM. In *4th International Conference on Sustainable Expert Systems* (pp. 826–830). IEEE. https://doi.org/10.1109/ICSES63445.2024.10762963

Nor, S. R. M., Husin, M. R., Jaeb, M. Z. M., & Naing, N. N. (2021). Socio-demographic profile and prevalence of tuberculosis (TB) treatment outcomes among tuberculosis/human immunodeficiency virus (TB/HIV) co-infected patients in Kelantan. *Pertanika Journal of Science and Technology*, *29*(4), 2279–2293. https://doi.org/10.47836/PJST.29.4.03

Pai, M., Behr, M. A., Dowdy, D., Dheda, K., Divangahi, M., Boehme, C. C., Ginsberg, A., Swaminathan, S., Spigelman, M., Getahun, H., Menzies, D., & Raviglione, M. (2016). Tuberculosis. *Nature Reviews Disease Primers*, *2*, 16076. https://doi.org/10.1038/nrdp.2016.76

Panicker, R. O., Kalmady, K. S., Rajan, J., & Sabu, M. K. (2018). Automatic detection of tuberculosis bacilli from microscopic sputum smear images using deep learning methods. *Biocybernetics and Biomedical Engineering*, *38*(3), 691–699. https://doi.org/10.1016/j.bbe.2018.05.007

Perez-Siguas, R., Matta-Solis, E., Remuzgo-Artezano, A., Matta-Solis, H., Matta-Perez, H., & Perez-Siguas, L. (2023). Chest X-ray imaging system for early detection of tuberculosis. In *Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies* (pp. 1–4). IEEE. https://doi.org/10.1109/ICAECT57570.2023.10117936

Rachmad, A., Chamidah, N., & Rulaningtyas, R. (2020). *Mycobacterium tuberculosis* images classification based on combining of convolutional neural network and support vector machine. *Communications in Mathematical Biology and Neuroscience*, *2020*, 85. https://doi.org/10.28919/cmbn/5035

Rachmad, A., Syarief, M., Hutagalung, J., Hernawati, S., Rochman, E. M. S., & Asmara, Y. P. (2024). Comparison of CNN architectures for *Mycobacterium tuberculosis* classification in sputum images. *Ingénierie des Systèmes d'Information*, *29*(1), 49–56. https://doi.org/10.18280/isi.290106

Rahman, T., Khandakar, A., Kadir, M. A., Islam, K. R., Islam, K. F., Mazhar, R., Hamid, T., Islam, M. T., Kashem, S., Mahbub, Z. B., Ayari, M. A., & Chowdhury, M. E. H. (2020). Reliable tuberculosis detection

using chest X-ray with deep learning, segmentation and visualization. *IEEE Access*, *8*, 191586–191601. https://doi.org/10.1109/ACCESS.2020.3031384

Rajaraman, S., & Antani, S. K. (2020). Modality-specific deep learning model ensembles toward improving TB detection in chest radiographs. *IEEE Access*, *8*, 27318–27326. https://doi.org/10.1109/ACCESS.2020.2971257

Rajaraman, S., Folio, L. R., Dimperio, J., Alderson, P. O., & Antani, S. K. (2021). Improved semantic segmentation of tuberculosis - Consistent findings in chest X-rays using augmented training of modality-specific U-Net models with weak localizations. *Diagnostics*, *11*(4), 616. https://doi.org/10.3390/diagnostics11040616

Rajaraman, S., Zamzmi, G., Folio, L. R., & Antani, S. (2022). Detecting tuberculosis-consistent findings in lateral chest X-rays using an ensemble of CNNs and vision transformers. *Frontiers in Genetics*, *13*, 864724. https://doi.org/10.3389/fgene.2022.864724

Riza, B. S., Na'am, J., & Sumijan, S. (2022). Tuberculosis extra pulmonary bacilli detection system based on Ziehl Neelsen images with segmentation. *MATRIK*, *22*(1), 139–148. https://doi.org/10.30812/matrik.v22i1.1939

Saini, A., Guleria, K., & Sharma, S. (2023). Deep learning-based model for the classification of tuberculosis using convolution neural network. In *4th IEEE Global Conference for Advancement in Technology* (pp. 1–6). IEEE. https://doi.org/10.1109/GCAT59970.2023.10353329

Samuel, D. J., & Baskaran, R. K. (2021). *Design to automate the detection and counting of tuberculosis (TB) bacilli*. arXiv. https://doi.org/10.48550/arXiv.2105.11432

Sarawagi, K., Pagrotra, A., Dhiman, H., & Singh, N. (2024). Self-trained convolutional neural network (CNN) for tuberculosis diagnosis in medical imaging. *Cureus*, *16*(6), e63356. https://doi.org/10.7759/cureus.63356

Shah, M. I., Mishra, S., Yadav, V. K., Chauhan, A., Sarkar, M., Sharma, S. K., & Rout, C. (2017). Ziehl-Neelsen sputum smear microscopy image database: A resource to facilitate automated bacilli detection for tuberculosis diagnosis. *Journal of Medical Imaging*, *4*(2), 027503. https://doi.org/10.1117/1.Jmi.4.2.027503

Shelomentseva, I. G., & Chentsov, S. V. (2021). Classification of microscopy image stained by Ziehl–Neelsen method using different architectures of convolutional neural network. In B. Kryzhanovsky, W. Dunin-Barkowski, V. Redko, & Y. Tiumentsev (Eds.), *Advances in Neural Computation, Machine Learning, and Cognitive Research IV* (pp. 26–275). Springer. https://doi.org/10.1007/978-3-030-60577-3_32

Shwetha, V., Prasad, K., Mukhopadhyay, C., Banerjee, B., & Chakrabarti, A. (2021). Automatic detection of bacilli bacteria from Ziehl-Neelsen sputum smear images. In *2nd International Conference on Communication, Computing and Industry 4.0* (pp. 1–5). IEEE. https://doi.org/10.1109/C2I454156.2021.9689283

Sua, L. F., Bolaños, J. E., Maya, J., Sánchez, A., Medina, G., Zúñiga-Restrepo, V., & Fernández-Trujillo, L. (2021). Detection of mycobacteria in paraffin-embedded Ziehl–Neelsen–stained tissues using digital pathology. *Tuberculosis*, *126*, 102025. https://doi.org/10.1016/j.tube.2020.102025

Suliman, Q., Md. Said, S., & Ying, L. P. (2019). Timing and prognostic factors of tuberculosis treatment interruption. *Pertanika Journal of Science and Technology*, *27*(4), 1509–1525.

Surani, C., Kumar, S., Chauhan, M., Chawda, H., Ramanuj, A. K., & Trivedi, K. (2021). Comparison between Ziehl-Neelsen staining and fluorescent staining of sputum samples to detect acid fast bacilli in suspected case of pulmonary tuberculosis at tertiary care hospital, Amreli, Gujarat. *Indian Journal of Microbiology Research*, *8*(4), 302–307. https://doi.org/10.18231/j.ijmr.2021.061

Tamura, G., Llano, G., Aristizábal, A., Valencia, J., Sua, L., & Fernandez, L. (2024). Machine-learning methods for detecting tuberculosis in Ziehl-Neelsen stained slides: A systematic literature review. *Intelligent Systems with Applications*, *22*, 200365. https://doi.org/10.1016/j.iswa.2024.200365

Tasci, E., Uluturk, C., & Ugur, A. (2021). A voting-based ensemble deep learning method focusing on image augmentation and preprocessing variations for tuberculosis detection. *Neural Computing and Applications*, *33*, 15541–15555. https://doi.org/10.1007/s00521-021-06177-2

Tiwari, M., Patankar, M., Chaurasia, V., Shandilya, M., Kumar, A., & Potnis, A. (2023). Detection of tuberculosis bacilli using deep learning. In *1st International Conference on Innovations in High Speed Communication and Signal Processing* (pp. 492–496). IEEE. https://doi.org/10.1109/IHCSP56702.2023.10127220

Toba, S., Mitani, Y., Yodoya, N., Ohashi, H., Sawada, H., Hayakawa, H., Hirayama, M., Futsuki, A., Yamamoto, N., Ito, H., Konuma, T., Shimpo, H., & Takao, M. (2020). Prediction of pulmonary to systemic flow ratio in patients with congenital heart disease using deep learning-based analysis of chest radiographs. *JAMA Cardiology*, *5*(4), 449–457. https://doi.org/10.1001/jamacardio.2019.5620

Tobin, E. H., & Tristram, D. (2024). *Tuberculosis overview*. StatPearls Publishing.

Tummalapalli, G., Prasad, C. B., & Kousik, J. (2024). Analysing the performance of Dense Net in the context of tuberculosis disease. In *International Conference on Advancements in Power, Communication and Intelligent Systems* (pp. 1–7). IEEE. https://doi.org/10.1109/APCI61480.2024.10617013

Witarto, A. B., Ceachi, B., Popp, C., Zurac, S., Daha, I. C., Sari, F. E., Putranto, N., Pratama, S., Octavianus, B. P., Nichita, L., Gerald Dcruz, J., Mogodici, C., Cioplea, M., Sticlaru, L., Busca, M., Stefan, O., Tudor, I., Dumitru, C., Vilaia, A., . . . Mustatea, P. (2024). AI-based analysis of Ziehl-Neelsen-stained sputum smears for *Mycobacterium tuberculosis* as a screening method for active tuberculosis. *Life*, *14*(11), 1418. https://doi.org/10.3390/life14111418

World Health Organization. (2024). *Global tuberculosis report 2024*. WHO. https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2024

Xie, Y., Wu, Z., Han, X., Wang, H., Wu, Y., Cui, L., Feng, J., Zhu, Z., & Chen, Z. (2020). Computer-aided system for the detection of multicategory pulmonary tuberculosis in radiographs. *Journal of Healthcare Engineering*, *2020*, 9205082. https://doi.org/10.1155/2020/9205082

Yang, M., Nurzynska, K., Walts, A. E., & Gertych, A. (2020). A CNN-based active learning framework to identify mycobacteria in digitized Ziehl-Neelsen stained human tissues. *Computerized Medical Imaging and Graphics*, *84*, 101752. https://doi.org/10.1016/j.compmedimag.2020.101752

Zaizen, Y., Kanahori, Y., Ishijima, S., Kitamura, Y., Yoon, H.-S., Ozasa, M., Mukae, H., Bychkov, A., Hoshino, T., & Fukuoka, J. (2022). 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), 709. https://doi.org/10.3390/diagnostics12030709

Zaporojan, N., Negrean, R. A., Hodișan, R., Zaporojan, C., Csep, A., & Zaha, D. C. (2024). Evolution of laboratory diagnosis of tuberculosis. *Clinics and Practice*, *14*(2), 388–416. https://doi.org/10.3390/clinpract14020030

Zurac, S., Mogodici, C., Poncu, T., Trăscău, M., Popp, C., Nichita, L., Cioplea, M., Ceachi, B., Sticlaru, L., Cioroianu, A., Busca, M., Stefan, O., Tudor, I., Voicu, A., Stanescu, D., Mustatea, P., Dumitru, C., & Bastian, A. (2022). A new artificial intelligence-based method for identifying *Mycobacterium tuberculosis* in Ziehl–Neelsen stain on tissue. *Diagnostics*, *12*(6), 1484. https://doi.org/10.3390/diagnostics12061484

Pertanika J. Sci. & Technol. 33 (6): 2465 - 2490 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Extraction and Characterisation of Natural Fibres from *Pennisetum purpureum*: Influence of Harvesting Age on Chemical, Mechanical, and Thermal Properties

Mohd Fairus Kayat[1], Ridhwan Jumaidin[1,2]*, Mastura Mohamad Taha[1] and Fahmi Asyadi Md Yusof[3]

[1]Fakulti Teknologi dan Kejuruteraan Industri dan Pembuatan, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

[2]Mechanical Engineering Programme, Faculty of Engineering, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia

[3]Malaysian Institute of Chemical and Bioengineering Technology, Taboh Naning, 78000 Alor Gajah, Melaka, Malaysia

ABSTRACT

This study examined new bio-based, natural cellulosic fibres from the *Pennisetum purpureum* plant, determining the optimal harvesting age at 30, 45, and 60 days for the first time. The fibres were extracted using a water-retting process and investigated for their potential as reinforcement materials in polymer composites. The characterisation process involved analysing the fibre's chemical composition, as well as its physical, thermal, mechanical, crystallinity, and morphological properties. Chemical composition was determined using standard chemical analysis, while mechanical properties were evaluated through tensile testing. X-ray diffraction (XRD) was used to assess the fibre's crystallinity, and thermogravimetric analysis (TGA) measured thermal stability. Additionally, scanning electron microscopy (SEM) was employed to study the fibre's surface morphology. The results indicated that mechanical properties were influenced by the ratio of cellulose to lignin. At 30 days, the fibre achieved optimal mechanical performance due to an elevated lignin concentration counteracting reduced cellulose levels (tensile strength: 67.54 MPa, modulus: 3.14 GPa, cellulose: 35.61%, lignin: 16.47%). This highlights that optimal fibre performance depends on the interaction of chemical constituents rather than cellulose content alone. XRD analysis revealed that the crystallinity index of *P. purpureum* fibres was 51.19, 49.40, and 50.56% for 30-, 45-, and 60-day fibres, respectively, with the highest value at 30 days, suggesting better molecular organisation due to reduced amorphous components. Thermogravimetric

ARTICLE INFO

Article history:
Received: 24 February 2025
Accepted: 29 April 2025
Published: 25 September 2025

DOI: https://doi.org/10.47836/pjst.33.6.04

E-mail addresses:
mfairus.kayat@gmail.com (Mohd Fairus Kayat)
ridhwanj@ums.edu.my (Ridhwan Jumaidin)
mastura.taha@utem.edu.my (Mastura Mohamad Taha)
fahmiasyadi@unikl.edu.my (Fahmi Asyadi Md Yusof)
* Corresponding author

© Universiti Putra Malaysia Press

analysis showed that fibre age affected thermal degradation, with older fibres exhibiting greater hemicellulose breakdown and improved thermal stability. The 45-day fibre had the lowest moisture loss (4.78%) and the highest char residue (28.54%), indicating better thermal resistance than the 30- and 60-day fibres. The main weight loss due to cellulose and hemicellulose decomposition occurred between 276–390°C, with the 60-day fibre showing a slightly higher peak degradation temperature (390°C), suggesting a more structured cellulose composition. SEM analysis revealed that the fibre surface was perforated and coarse, with a central lumen, further supporting its potential as a reinforcing material in thermoplastic green composites. This indicated that *P. purpureum* fibre has strong potential as a reinforcing material for thermoplastic green composites.

Keywords: Cellulose, crystallinity index, mechanical properties, *Pennisetum purpureum* fibre, scanning electron microscopy, thermogravimetric analysis, X-ray diffraction

INTRODUCTION

Future generations will confront substantial hazards, including global warming, excessive emissions, and the lack of natural resources (Kamaruddin et al., 2021). The extensive reliance on petroleum-based plastics in everyday life has led to serious environmental problems, especially when managing waste (Diyana, Jumaidin, Selamat, Ghazali, et al., 2021). These materials do not decompose readily in nature, whether they are deposited on land or in the ocean, resulting in significant issues for civilisation, wildlife, and the ecosystem. Consequently, the need for more ecologically sustainable materials is significantly rising to protect the environment and maintain ecosystem equilibrium. Finding eco-friendly materials made from renewable resources offers a sustainable alternative to synthetic materials. This approach helps conserve the planet's resources while reducing environmental impact. In recent decades, natural fibres have become increasingly popular due to their unique characteristics, including biodegradability, abundant availability, affordability, non-toxicity, low density, environmental friendliness, recyclability, and excellent balance of strength and weight (Jumaidin et al., 2020; Kamaruddin et al., 2021; Tarique et al., 2021). The increasing demand for traditional fibres, coupled with their limited production, has driven the exploration of new natural fibres made up of cellulose, hemicellulose, lignin, wax, ash, and other soluble compounds (Karimah, Ridho, Munawar, Adi, et al., 2021).

Cellulose is a type of linear polysaccharide made up of glucose molecules. Natural fibres with high cellulose content tend to be stronger and more stable (Tarique et al., 2021). To understand whether a new natural fibre is more crystalline or amorphous, researchers usually calculate its crystallinity index (CI). Natural fibres' properties - physical, mechanical, and chemical—can vary greatly depending on their chemical composition, how mature they are, the part of the plant they come from, the region where the plant was grown, soil nutrient levels, and climate conditions (Biradar et al., 2023).

The importance of natural fibres is highlighted through a comprehensive review of existing research in the same study. A prior study found that the fibre's cellulose levels and the index of crystallinity directly influenced their properties. Natural fibres are often incorporated into composite materials to improve their strength and overall performance (Khalid et al., 2021). Its cellulose content and CI greatly influence the effectiveness of a fibre. These factors play a key role in determining the fibre's performance. The balance between crystalline cellulose and the more amorphous components, like hemicellulose, lignin, and wax, varies depending on the plant's growth environment and conditions. Characteristics of polymer composites depend on various parameters, such as the type of resin, alignment with respect to the respective fibres, and strong bonding between the fibres and the matrix (Chakravarthy K et al., 2020).

Pennisetum purpureum, commonly referred to as Napier grass, is a vital fodder crop in humid tropical regions, valued for its exceptional dry matter production per unit area, surpassing many other grass species. It is extensively cultivated across tropical climates, particularly in Southeast Asia, where annual rainfall typically reaches around 1,000 mm (Onjai-uea et al., 2023). *Pennisetum purpureum* exhibits the ability to thrive in soil conditions that are unfavourable for numerous plant species (Purba & Paengkoum, 2019). *Pennisetum purpureum* exhibits vegetative propagation and possesses the ability to endure many cuttings while swiftly regenerating. This characteristic enables it to yield highly desirable fodder for cattle, particularly in terms of foliage content (Onjai-uea et al., 2023).

The plant's age typically determines the optimal timing for defoliation to ensure maximum production and quality, and recommended guidelines can vary accordingly. For maximum results and high-quality feed, it is recommended to defoliate the grass every 70 days or when the plant reaches a height of 120-150 cm (Chuwongpanich et al., 2019). In addition, it has been proposed by the researcher that grass should be defoliated between 60 and 85 days of age (Ferreira et al., 2019). However, it has also been noted that the harvesting period of the grass can be done after 25 to 30 days during the wet season, or 50 to 60 days during the dry season (Onjai-uea et al., 2023). Meanwhile, increasing the cutting interval to 45 to 60 days will lead to a greater amount of dry matter produced compared to cutting every 30 days (Zaini et al., 2021). The investigations found that as the plant matured, it produced more dry matter, but the overall quality of the plant declined. It indicated that the different harvesting ages of *P. purpureum* resulted in notable variations in the chemical properties of the grass biomass. Grass harvested at a later age contained higher levels of the components that are difficult to degrade, such as lignin. Both the initial features and degradability were altered using a separate method (Chanpla et al., 2018).

A study explored natural fibres, focusing on their thermal, mechanical, and morphological properties, along with their chemical composition (Imraan et al., 2023; Sunny & Rajan, 2022). The impact of betel nut husk (BNH) at different stages of maturity

(raw, ripe, and fully matured) on its mechanical, physical, thermal, and morphological properties was investigated (Sunny & Rajan, 2022). The study revealed that fibres from the ripe stage had the highest tensile strength. On the other hand, sugar palm fibres have been characterised in terms of their tensile and thermal properties, where the maximum tensile strength is found in natural fibre comparison, which attained an ideal chemical composition belonging to that category, featuring a high concentration of cellulose, along with hemicellulose and lignin (Imraan et al., 2023).

While research has explored lignocellulosic fibres as strengthening agents in composite materials, no studies have specifically examined natural cellulosic fibres and the properties of *P. purpureum* fibre based on their optimal harvesting age. This study aimed to analyse the chemical, physical, mechanical, and thermal properties of *P. purpureum* at different fibre ages, comparing them to other well-known natural fibres. The results aim to enhance our understanding of how this natural fibre can be used in various biodegradable products. Despite its potential, there has been limited research on the use and applications of *P. purpureum* fibre and its composites. This study examined the fibre's properties through tests measuring density, chemical composition, mechanical strength, thermal stability, morphology, and functional groups. The findings were then contrasted with other natural fibres.

EXPERIMENTAL METHODS

Raw Materials Preparation

The *P. purpureum grass* was sourced from Ladang Napier Selatan in Rembau, Negeri Sembilan, Malaysia. The fibres were chosen based on the three typical harvesting ages of *P. purpureum grass*, which were 30, 45, and 60 days (Onjai-uea et al., 2023; Zaini et al., 2021).

Fibres Preparation

The surface of the harvested *P. purpureum* grass was washed with clean water to remove any dirt or unwanted particles. The cleaned stems were submerged in fresh water for 14 days to undergo retting, a process that uses microbial activity to break down the plant material. The tank's materials were regularly assessed to confirm that fibres were sufficiently loosened for extraction. After the retting process, the fibres were carefully extracted from the stems by hand. The extracted fibres were rinsed with distilled water to remove any dust, then sun-dried for 48 hr. Afterwards, the fibres were oven-dried at 85°C for an additional 24 hr to eliminate any remaining moisture (Huzaifah et al., 2017). The dried stems were subsequently split into smaller segments prior to the grinding procedure that yielded *P. purpureum* fibre. The water retting flow mechanisms are illustrated in Figures 1(a) to (f).

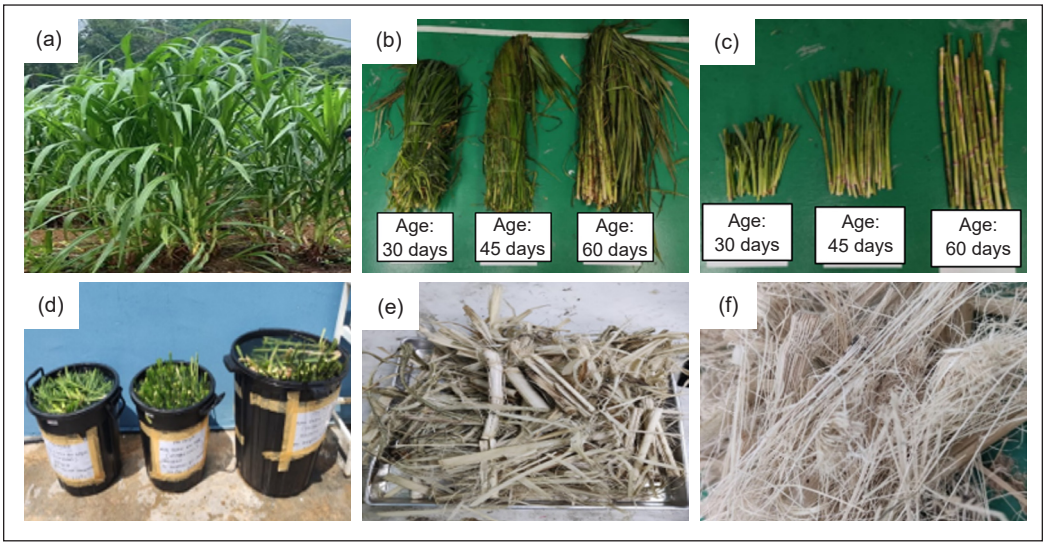


Figure 1. (a) *Pennisetum purpureum* plant; (b) *Pennisetum purpureum* selected at three harvesting ages: 30, 45, and 60 days; (c) cleaned stem; (d) water retting of *Pennisetum purpureum*; (e) dried *Pennisetum purpureum*; and (f) extracted fibres

Chemical Composition

The chemical composition of *P. purpureum* fibre was assessed through neutral detergent fibre (NDF), acid detergent fibre (ADF), and acid detergent lignin (ADL) in accordance with the ISO 13906:2008 standard. The calculated weight percentages of cellulose and hemicellulose were determined using Equations 1 and 2, respectively:

$$\text{Cellulose} = \text{ADF} - \text{ADL} \quad [1]$$

$$\text{Hemicellulose} = \text{NDF} - \text{ADF} \quad [2]$$

Physical Properties

Diameter

The diameter of *P. purpureum* fibres was measured using a Zeiss optical microscope (Axiovert 200, Carl Zeiss Light Microscopy, Germany), analysing 15 individual fibre samples (Kamaruddin et al., 2021). The average diameter was calculated by measuring three different points on each photograph and calculating the mean value (Razali et al., 2015).

Density

The fibre density of *P. purpureum* was determined using an AccuPyc 1340 TEC pycnometer (Micromeritics Instrument Corporation, USA). The pycnometer was used to determine the volume and density of powdered fibre, following ASTM D792 (ASTM International,

2020b), as outlined by Kamaruddin et al. (2021). Helium gas was employed at a constant temperature for this experiment. For each sample, three individual measurements were taken, and the average value was calculated for the data analysis (Ilyas et al., 2017).

Moisture Content

Moisture content was evaluated using five samples, which were dried in an oven at 105°C for 24 hr. To determine the moisture content, the initial weight before drying (Mi, g) and the final weight after drying (Mf, g) were recorded (Diyana, Jumaidin, Selamat, Alamjuri, et al., 2021; Ilyas et al., 2017; Kamaruddin et al., 2021). The moisture content of *P. purpureum* fibre was determined using Equation 3:

$$\text{Moisture content} = \frac{\text{Mi} - \text{Mf}}{\text{Mi}} \times 100\% \qquad [3]$$

Tensile Properties – Single Tensile Test

The ASTM C1557 (ASTM International, 2020a) guideline was utilised for assessing the tensile qualities of *P. purpureum* fibre using a universal testing machine equipped with a 1 kN load cell (Engelbrecht-Wiggans & Forster, 2023). Experiments were conducted with a 25 mm gauge length and a crosshead speed of 1 mm/min. *Pennisetum purpureum* fibre was meticulously selected using an optical microscope before testing to ensure the fibres were undamaged (Razali et al., 2015). Each fibre was secured to the sample holder, and 15 samples were utilised for the assessment of tensile properties, as shown in Figure 2.

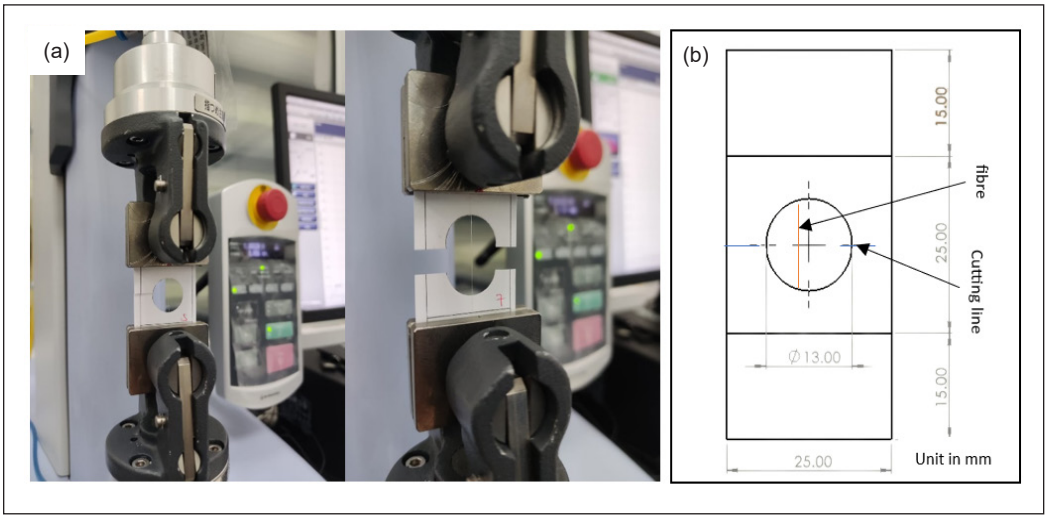


Figure 2. (a) Preparation of samples for tensile testing; and (b)Top-view arrangement of the sample during tensile testing

Thermal Characterisation

Thermogravimetric Analysis (TGA)

To evaluate the thermal stability of *P. purpureum* fibre, a thermogravimetric analyser from TA Instruments (Mettler-Toledo AG, Switzerland) was used, as part of the Thermal Analysis Q series, to determine the fibres' potential for high-temperature applications. A measured amount of *P. purpureum* fibres was placed in an alumina crucible and positioned in a heating chamber. The study was performed inside a nitrogen chamber at a temperature control of 25°C and gradually increased to 900°C at a steady and regular rate of 10°C per minute (Kamaruddin et al., 2021).

Scanning Electron Microscopy (SEM)

The cross-section and morphology of *P. purpureum* fibres were analysed using a Hitachi S-3400N scanning electron microscope (Japan) at an acceleration voltage of 15 kV. Before analysis, the fibres were coated with a thin layer of gold to enhance the imaging. The specimens were 50 mm in length, and the results were obtained through imaging at various magnification levels (Arul Marcel Moshi et al., 2020).

Fourier Transform Infrared Spectroscopy (FTIR)

FTIR spectroscopy was used to analyse all samples and identify the functional groups present in *P. purpureum* fibre. The analysis was carried out with an IR spectrometer (Nicolet 6700 AEM, Thermo Nicolet Corporation, USA). Fine powders were mixed with potassium bromide (KBr, Mody Chemi-Pharma Limited, India) and ground into a fine powder. The powdered sample of approximately 2 mg was combined with the KBr to make a 1-mm-thick disc. The FTIR spectra were recorded from 4000 to 500 cm^{-1} (Diyana, Jumaidin, Selamat, Alamjuri, et al., 2021).

XRD Analysis

The XRD method was selected to assess the CI and crystalline size of *P. purpureum* fibre, utilising a Rigaku D/max 2500 X-ray powder diffractometer (Rigaku, Japan) operating with copper (Cu) radiation at 40 kV and 30 mA, with a wavelength of 0.15406 nm. A scanning rate of 2°/min was employed to examine the samples within the diffraction angle range of 10 to 40° at room temperature. The CI of *P. purpureum* fibre was calculated using the Segal equation, as stated in Equation 4 (Ilyas et al., 2018).

$$CI = \frac{I_{002} - I_{am}}{I_{002}} \times 100\% \qquad [4]$$

Peak intensities for the amorphous and crystalline components are denoted as I002 and Iam, respectively. The crystallite size (CS) was calculated with the use of Scherrer's formula as shown in Equation 5.

$$CS = \frac{k\lambda}{\beta \cos\theta} \times 100\% \quad [5]$$

In this equation, k = 0.89 is Scherrer's constant, λ = 0.1541 nm is the wavelength of the radiation, β represents the full width at half-maximum of the peak in radians, and θ is the corresponding Bragg angle.

RESULTS AND DISCUSSION

Chemical Composition

Natural fibre consists of three key components: cellulose, hemicellulose, and lignin. Cellulose serves as the main structural framework of the fibre, while hemicellulose and lignin surround the cellulose regions, contributing to the fibre's strength, stiffness, and overall stability (Kabir et al., 2012). Analysing the chemical composition of natural fibres is essential for assessing their mechanical characteristic, especially in strength, as the concentration of different chemical components on the fibre's surface plays a significant role in determining their strength (Arul Marcel Moshi et al., 2020). Table 1 compares the results from the examination of the chemical composition of *P. purpureum* fibre across three distinct fibre ages. The highest cellulose content (approximately 45.9%) was observed at 45 days. Meanwhile, the lowest cellulose content was at 30 days (35.61%). At 60 days, cellulose stated the mid value between 30 and 60 days (38.82%). Hemicellulose remained relatively stable across all ages, with a slight increment from 25.03% at 45 days to 27.38% at 60 days. Lignin content, however, dropped significantly at 45 days (10.08%) from 16.47% at 30 days and rose to 13% by 60 days. Based on current observation, the data highlighted 45 days as the optimal harvesting age for balancing high cellulose and low lignin content. The results were then compared to those of other natural fibres, as shown in Table 1.

At 45 days of age, *P. purpureum* fibres contained 45.90% cellulose, relatively higher compared to other natural fibres, such as *Axonopus compressus* (27.28%) (Wahid et al., 2023), *Hylocereus polyrhizus* (34.45) (Taharuddin et al., 2023), *Cymbopogon citratus* leaves (37.56%) (Kamaruddin et al., 2023), and *Arenga pinnata* (44.53%) (Huzaifah et al., 2017). Meanwhile, the cellulose content for 30 and 60 days was lower than that of the other fibres mentioned. The cellulose content for all fibre ages of *P. purpureum* was lower than that of *Pandanus amaryllifolius* (48.79%) (Diyana, Jumaidin, Selamat, Alamjuri, et al., 2021), kenaf (50-55%) (Abdullah et al., 2020), ramie (68-76%) (Goda et al., 2006; Khoo et al., in press; Maideliza et al., 2017), and roselle (60-64%) (Razali et al., 2015).

Table 1
Comparison of the chemical composition of Pennisetum purpureum *fibre and other natural fibres*

Fibre	Fibre age	Cellulose (% wt/wt)	Hemicellulose (% wt/wt)	Lignin (% wt/wt)	Ash (% wt/wt)	References
Pennisetum purpureum	30-days	35.61	26.63	16.47	2.29	Current study
	45-days	45.90	25.03	10.08	2.61	
	60-days	38.82	27.38	13.00	3.64	
Kenaf	12-weeks	50.25-55.43	32.29-37.73	14.87-17.99	0.38-2.45	Abdullah et al. (2020)
	16-weeks	50.85-56.10	24.14-35.41	14.23-17.97	0.29-1.84	
Ramie	50-days	68.60-76.20	13.10-16.70	0.60-0.70	-	Goda et al. (2006); Khoo et al. (in press); Maideliza et al. (2017)
	60-days					
	70-days					
Roselle	3-months	64.50	20.23	10.26	1.25	Razali et al. (2015)
	6-months	60.51	16.27	10.26	1.03	
	9-months	58.63	20.82	7.87	2.08	
Axonopus compressus	-	27.28	29.29	11.14	-	Wahid et al. (2023)
Hylocereus polyrhizus		34.45	-	6.73	17.40	Taharuddin et al. (2023)
Pandanus amaryllifolius	-	48.79	19.95	18.64	-	Diyana, Jumaidin, Selamat, Alamjuri, et al. (2021)
Cymbopogon citratus leaves	-	37.56	29.29	11.14	4.28	Kamaruddin et al. (2021)
Arenga pinnata	-	44.53	10.01	41.97	0.955	Huzaifah et al. (2017)

The elevated cellulose content in the fibre enhanced the mechanical properties (Sheltami et al., 2012). The results in Table 1 show that the cellulose content of *P. purpureum* fibre varied significantly with age, which may account for the differing tensile values observed in mechanical testing, in addition to the impact of fibre diameter. The presence of a considerable amount of amorphous compounds in *P. purpureum* fibre, including hemicellulose, lignin, ash, and waxes, may also exert an influence (Campos et al., 2018).

Physical Properties

Diameter and Density

Table 2 presents the values acquired for *P. purpureum* fibre's diameter. Measuring the diameter is crucial for assessing the fibre's mechanical qualities, particularly tensile strength. The diameter of the *P. purpureum* fibre was measured with an optical microscope, as shown in Figure 3. The average diameters of the fibre over 30, 45, and 60 days were determined to be 68, 138, and 162 µm, respectively, equivalent to those of kenaf (57.8-67.5 µm) (Abdullah et al., 2020), roselle (40-10, 90-120, and 90-150 µm) (Razali et al., 2015), *H. polyrhizus* (0.45 µm)

(Taharuddin et al., 2023), *P. amaryllifolius* (368.57 ± 50.47 µm) (Diyana, Jumaidin, Selamat, Alamjuri, et al., 2021), *C. citratus* (326.67 ± 45.77 µm) (Kamaruddin et al., 2021), and *A. pinnata* (0.40 ± 0.08 µm) (Huzaifah et al., 2017). The physical properties are primarily affected by factors like the fibre's origin, the plant's condition, its age, and the method of extraction (Reddy & Yang, 2005).

Density is a crucial parameter in determining the mass of natural fibres and plays a key role in the overall mass of composite materials made from these fibres. The densities of *P. purpureum* fibre at 30, 45, and 60 days were quantified using a pycnometer (Micromeritics Instrument Corporation, USA) with recorded values of 0.67, 1.03, and 1.23 g/cm³, respectively. These values were comparatively lower than those of other natural fibres, such as ramie (1.44, 1.45, and 1.46 g/cm³) (Goda et al., 2006; Khoo et al., in press;

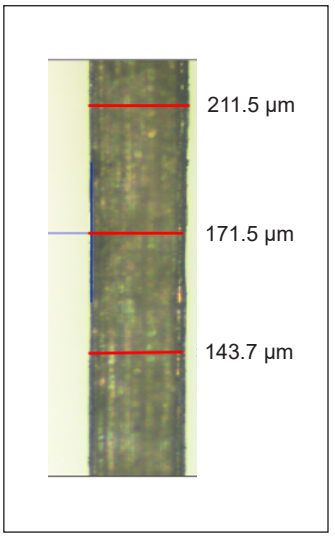


Figure 3. Optical microscopic view of *Pennisetum purpureum* fibre at 100× magnification

Table 2
Comparison of Pennisetum purpureum *fibre's diameter and density with other various fibres*

Fibre	Fibre age	Diameter (µm)	Density (g/cm³)	Moisture content (%)	References
Pennisetum purpureum	30-day	78	0.67	11.06	Current study
	45-day	146	1.03	11.84	
	60-day	159	1.23	11.92	
Kenaf	12-week	57.80-67.50	-	8.33-10.02	Abdullah et al. (2020)
	16-week				
Ramie	50-day	-	1.44	-	Goda et al. (2006); Khoo et al. (in press); Maideliza et al. (2017)
	60-day	-	1.45	-	
	70-day	-	1.46	-	
Roselle	3-month	40-100	1.33	5.80	Razali et al. (2015)
	6-month	80-120	1.42	4.90	
	9-month	90-150	1.42	3.70	
Axonopus compressus	-	-	-	14.95	Wahid et al. (2023)
Hylocereus polyrhizus	-	590	0.45	9.70	Taharuddin et al. (2023)
Pandanus amaryllifolius	-	368.57 ± 50.47	-	6.00 ± 0.13	Diyana, Jumaidin, Selamat, Alamjuri, et al. (2021)
Cymbopogon citratus	-	326.67 ± 45.77	0.25 ± 0.002	5.20 ± 2.28	Kamaruddin et al. (2021)
Arenga pinnata	-	0.40 ± 0.08	1.46 ± 0.012	6.45 ± 1.07	Huzaifah et al. (2017)

Maideliza et al., 2017), roselle (1.44, 1.45, and 1.46 g/cm³) (Razali et al., 2015), *H. polyrhizus* (0.67 g/cm³) (Taharuddin et al., 2023), *C. citratus* (0.25 ± 0.002 g/cm³) (Kamaruddin et al., 2021), and *A. pinnata* (1.46 ± 0.012 g/cm³) (Huzaifah et al., 2017). Table 2 compares the *P. purpureum* fibre's diameter and density with other various fibres. The density of *P. purpureum* fibre highlights its potential for use in the production of low-weight composite materials.

Moisture Content

Moisture content is an essential factor in evaluating a new material for potential reinforcement in composite materials (Kamaruddin et al., 2021). Fibres with low moisture content form a stronger bond with the polymer matrix in composite materials (Jayaramudu et al., 2010). As shown in Table 3, *P. purpureum* fibre of 30, 45, and 60 days age recorded the moisture contents of 11.06, 11.84, and 11.92%, respectively, slightly lower compared to *A. compressus* of 14.95% (Wahid et al., 2023) and slightly higher to the other fibres – kenaf (8.33-10.02%) (Abdullah et al., 2020), roselle (5.80 to 3.70%) (Razali et al., 2015), *H. polyrhizus* (9.70%) (Taharuddin et al., 2023), *P. amaryllifolius* (6.00 ± 0.13%) (Diyana, Jumaidin, Selamat, Alamjuri, et al., 2021), *C. citratus* (5.20 ± 2.28) (Kamaruddin et al., 2021), and *A. pinnata* (6.45 ± 1.07) (Huzaifah et al., 2017). Nevertheless, elevated moisture content demonstrated considerable hydrophilic characteristics. This might show its chemical composition that increased its water adhesivity, making it potentially suitable for high moisture retention applications (Wahid et al., 2023). In summary, the moisture content patterns of *P. purpureum* fibre, in comparison to other natural fibres, underscore its low water retention capacity. This renders it adaptable for medium-stability applications while guaranteeing sufficient mechanical performance. Additional optimisation, including surface treatment, may enhance its functionality in moisture-sensitive settings, expanding its application possibilities.

Tensile Properties

Table 3 presents a comparison of the tensile properties of *P. purpureum* fibre based on fibre age: 30, 45, and 60 days. Maximum strength, which normalises the breaking force relative to the cross-sectional area, showed a different trend. The highest value appeared at 30 days (67.54 MPa), reflecting the fibre's strong tensile capability despite its small diameter. Strength declined significantly to 29.83 MPa at 45 days before rebounding slightly to 41.01 MPa at 60 days. The modulus of elasticity, or stiffness, declined with age, albeit non-linearly. The 30-day fibre exhibited the highest stiffness at 3.14 GPa, indicating it was the least prone to deformation under stress (Zor et al., 2023). By 45 days, the modulus dropped to 1.58 GPa, showing a significant reduction in stiffness, which might be attributed to changes in microfibril angle or structural alignment within the fibre (B. Zhang et al., 2020; Suryanto et al., 2014). At 60 days, the modulus increased to 2.18 GPa, suggesting some recovery of stiffness as the fibre matured (Gallos et al., 2017).

Table 3 also compares the results of tensile properties, strength, and modulus of different ages of *P. purpureum* fibre to those of other well-known natural fibres. The tensile strength of *P. purpureum fibre* depends on factors such as plant age, origin, fibre extraction method, and microstructure. Fibre failure can occur due to cracks forming and spreading from larger imperfections (Bezazi et al., 2014). Therefore, testing at least 15 replicates of each fibre sample is crucial to accurately determine its tensile properties. The remarkable tensile strength of natural fibres is primarily attributed to their high cellulose content and CI (Gurukarthik Babu et al., 2019). The tensile strength of *P. purpureum* fibre was lower compared to other natural fibres, such as kenaf (421–424 MPa) (Abdullah et al., 2020), ramie (859.15–1,150.06 MPa) (Goda et al., 2006; Khoo et al., in press; Maideliza et al., 2017), and Roselle (228.57-453.48 MPa) (Razali et al., 2015). Meanwhile, *P. purpureum* fibre demonstrated a higher tensile strength compared to aerial roots of the banyan tree (19.37 MPa) (Ganapathy et al., 2019), oil palm fruit (49 MPa) (Arul Marcel Moshi et al., 2020), and coconut tree leaf sheath (46.40 MPa) (Kulandaivel et al., 2020).

Table 3
Comparison of tensile properties between Pennisetum purpureum *fibre and various natural fibres*

Fibre	Fibre age	Tensile strength (MPa)	Tensile modulus (GPa)	Elongation at break (%)	References
Pennisetum purpureum	30-day	67.54	3.14	3.46	Current study
	45-day	29.83	1.58	3.38	
	60-day	41.01	2.18	3.03	
Kenaf	12-week	421	-	-	Abdullah et al. (2020)
	16-week	424	-	-	
Ramie	50-day	1,150.06	66.49	1.82	Goda et al. (2006); Khoo et al. (in press); Maideliza et al. (2017)
	60-day	1,089.95	± 57.00	± 1.40	
	70-day	859.15	54.56	1.36	
Roselle	3-month	453.48	-	-	Razali et al. (2015)
	6-month	247.28	-	-	
	9-month	228.57	-	-	
Pandanus amaryllifolius	-	45.61 ± 16.09	0.41 ± 0.18	8.17 ± 3.84	Diyana, Jumaidin, Selamat, Alamjuri, et al. (2021)
Cymbopogon citratus	-	43.81 ± 15.27	1.05 ± 0.33	0.84 ± 0.28	Kamaruddin et al. (2021)
Arenga pinnata	-	233.00 ± 71.17	4.19 ± 1.61	20.6 ± 9.29	Huzaifah et al. (2017)
Aerial roots of the banyan tree	-	19.37	1.80	1.80	Ganapathy et al. (2019)
Coconut tree leaf sheath	-	46.40	2.30	2.84	Kulandaivel et al. (2020)
Oil palm fruit	-	49	2.76	7	Arul Marcel Moshi et al. (2020)

Numerous studies have demonstrated that higher cellulose content and crystallinity typically correlate with greater tensile strength (Kamaruddin et al., 2021; Todkar & Patil, 2019). However, an excessive cellulose content can negatively impact tensile properties due to the inherently hydrophilic nature of cellulose. To further enhance mechanical strength with increased fibre loading, the use of a compatibiliser is recommended (Todkar & Patil, 2019). In general, from the results, despite having the lowest cellulose content (35.61%), the *P. purpureum* fibre at 30 days exhibited the highest tensile strength (67.54 MPa) and tensile modulus (3.14 GPa), likely due to its elevated lignin content (16.47%). Lignin provides structural reinforcement, contributing to the fibre's stiffness and strength (Karimah, Ridho, Munawar, Ismadi, et al., 2021). At 45 days, although the fibre reached its peak cellulose content (45.90%), it showed the lowest tensile strength (29.83 MPa) and modulus (1.58 GPa). This decline in mechanical properties was likely due to a significant reduction in lignin content (10.08%), which played a crucial role in binding cellulose fibrils and providing rigidity (Dorez et al., 2014; Oladele et al., 2020; Sultana, 2014). By 60 days, while cellulose content decreased to 38.82%, both tensile strength and modulus improved compared to 45 days. This improvement correlated to the increase in lignin content (13.00%), indicating that lignin compensated for the lower cellulose levels by enhancing the fibre's rigidity and resistance to deformation (Zor et al., 2023).

The mechanical properties of *P. purpureum* fibres are affected by the intricate relationship between their chemical composition and microstructural features. Although cellulose is essential in influencing tensile strength and modulus, the findings suggested that lignin content and its interplay with cellulose are of similar significance. This has underscored the necessity for a comprehensive approach to assessing fibre performance, considering both chemical composition and structural configuration. Subsequent research may concentrate on analysing cellulose crystallinity and fibre morphology to enhance comprehension of these correlations.

Thermogravimetric Analysis (TGA)

Natural fibres exhibit restricted thermal stability at elevated temperatures, which poses a challenge to their use as reinforcing materials in structural composites (Kamaruddin et al., 2021). Figure 4 illustrates the TGA and derivative thermogravimetric (DTG) curves for *P. purpureum* fibres at three different ages. This thermal analysis evaluates weight loss as the material undergoes heating. As shown in Table 4, three distinct phases of degradation were evident: an initial minor weight loss within 30–102°C, a significant weight loss between 276–390°C, and a final degradation stage extending to the maximum temperature.

The initial phase took place with weight losses of 10.89, 4.78, and 9.29% for fibres aged 30, 45, and 60 days, respectively. This phase was associated with the release of absorbed moisture, a characteristic of the hydrophilic properties of lignocellulosic materials

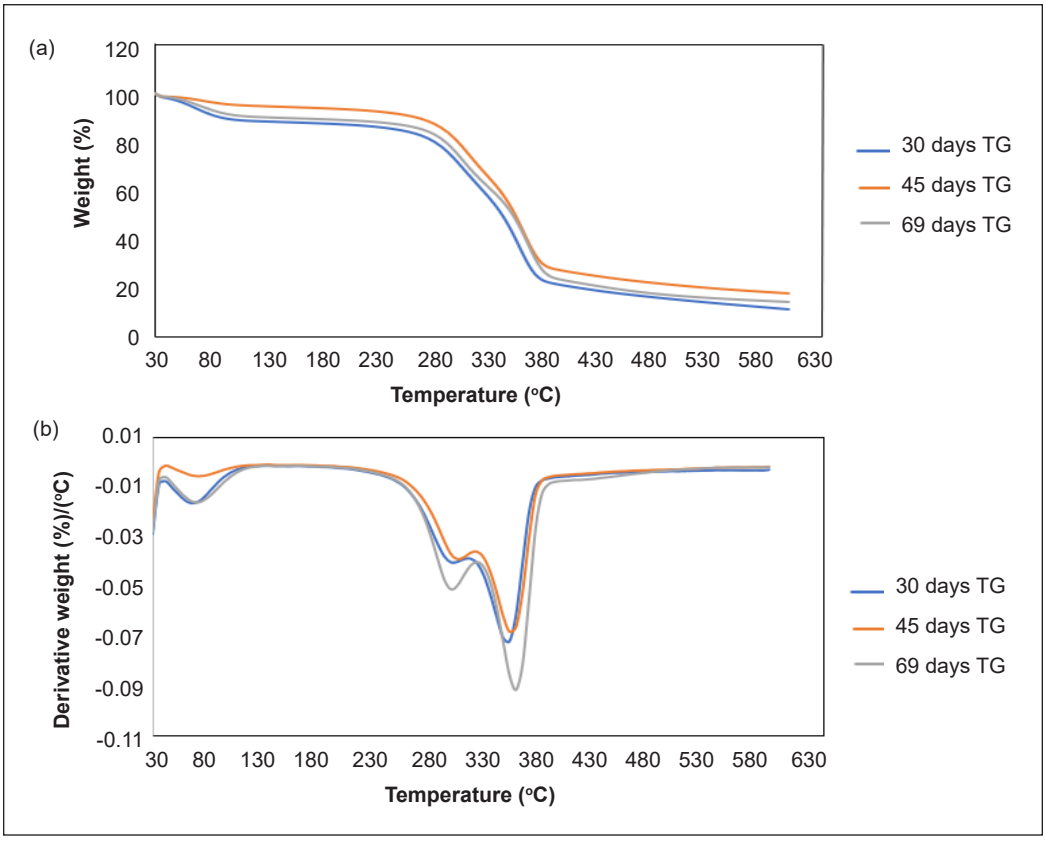


Figure 4. Thermogravimetric (TG) and derivative thermogravimetric (DTG) curves for *Pennisetum purpureum* fibres at various ages: (a) TG curve; and (b) DTG curve

(Boumediri et al., 2019; Mohammed et al., 2023). Subsequently, the fibres demonstrated thermal stability up to around 270°C, as indicated by the absence of notable peaks in both the TGA and DTG curves, which aligned with previous studies (Maache et al., 2017).

The second phase marked a significant weight loss—55.77, 58.40, and 58.99% for fibres aged 30, 45, and 60 days, respectively—resulting from the decomposition of cellulose and hemicellulose. This degradation occurred between 276–378°C, with DTG peak temperatures observed at 378°C for the 30- and 45-day fibres, and at 390°C for the 60-day fibre. This shift in peak temperature suggested the thermal decomposition of cellulose I and the completion of α-cellulose degradation (Abera et al., 2020).

The final degradation phase involved the fragmentation of waxy substances like lignin between 420–500°C, resulting in minor weight losses of 9.87, 8.28, and 7.54% for the respective fibre ages. At the end of the thermal process (600°C), char residue amounted to 23.47, 28.54, and 24.18% for the 30-, 45-, and 60-day fibres, respectively. The char residue primarily consists of carbon or carbon-based materials that resist further thermal

decomposition and remain as residue. Additionally, the presence of inorganic components may contribute to this residue, enabling quantitative estimations of fibre composition (Diyana, Jumaidin, Selamat, Alamjuri, et al., 2021). According to Boumediri et al. (2019), the char residue for *Juncus effusus* L. fibres was found to be 3.64%, while date palm fibres had a char residue of 18.29%. Table 4 shows the thermal degradation of *P. purpureum* fibre at different ages. The thermal stability of *P. purpureum* was comparable with other well-known natural fibres, as shown in Table 5.

Table 4
Thermal degradation properties of Pennisetum purpureum *fibre at different ages*

Sample age (day)	1st thermal degradation			2nd thermal degradation			3rd thermal degradation			Char residue (wt%)
	*T*1 (°C)	Weight loss (%)	*T*peak (°C)	*T*2 (°C)	Weight loss (%)	*T*peak (°C)	*T*3 (°C)	Weight loss (%)	*T*peak (°C)	
30	30 to 102	10.89	102	276 to 378	55.77	378	378 to 540	9.87	504	23.47
45	36 to 102	4.78	84	276 to 378	58.40	378	378 to 480	8.28	492	28.54
60	30 to 102	9.29	102	282 to 390	58.99	390	390 to 510	7.54	510	24.18

Table 5
Degradation temperature of various natural fibres

Natural fibre	Decomposition temperature, initial (°C)	Decomposition temperature, maximum (°C)	References
Pennisetum purpureum fibre – age 30 days	276	378	Current study
Pennisetum purpureum fibre – age 45 days	276	378	Current study
Pennisetum purpureum fibre – age 60 days	282	390	Current study
Axonopus compressus	228	360	Wahid et al. (2023)
Hylocereus polyrhizus	204	306	Taharuddin et al. (2023)
Pandanus amaryllifolius	220	370	Diyana, Jumaidin, Selamat, Alamjuri, et al. (2021)
Cymbopogon citratus	230	338	Kamaruddin et al. (2023)
Sugar palm	228	312	Huzaifah et al. (2017)

Morphological Properties

The morphology of natural fibres is a crucial aspect affecting their mechanical and physical properties, determining their appropriateness as reinforcing materials for composites (Kamaruddin et al., 2021). The SEM pictures illustrate the surface morphology of *P. purpureum* fibres at three distinct growth stages. Figure 5 shows the data for (a) 30 days, (b) 45 days, and (c) 60 days.

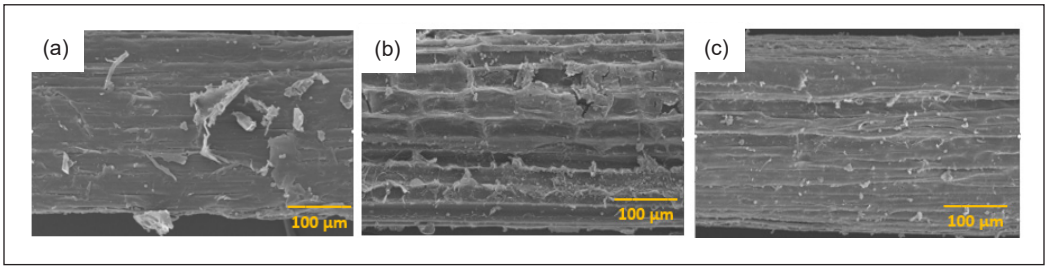


Figure 5. SEM micrographs of *Pennisetum purpureum* raw fibres at different ages: (a) 30 days; (b) 45 days; and (c) 60 days, recorded at 500× magnification with a 100 µm scale bar

At 30 days of fibre age, the surface exhibited a rough texture with several protruding fragments and loose particles. The visible flake-like features suggested incomplete formation or weak bonding of the cell wall layers at this early growth stage. These irregularities were likely due to the relatively high moisture content and the presence of non-fibrous elements, like hemicellulose and pectin, which were still abundant (Radakisnin et al., 2020). The 45-day fibre age showed a more compact and layered structure. The fibre's surface appeared to have fewer fragments than the 30-day sample, indicating ongoing cell wall maturation and deposition of lignin and cellulose (Radakisnin et al., 2020; W. Zhang et al., 2020; Yuan et al., 2024). The 60-day fibre age demonstrated the smoothest and most well-structured morphology. The layers were densely packed with minimal visible defects or loose debris (Radakisnin et al., 2020; Yuan et al., 2024). This aligned with the findings of Diyana, Jumaidin, Selamat, Alamjuri, et al. (2021) and Donaldson et al. (2016), which may be ascribed to the architecture consisting of microfibrils that were tightly arranged and interconnected by lignin, pectin, and other non-cellulosic substances. The architecture of microfibrils was observable and further substantiated the existence of lignin, which bound cellulose microfibrils and safeguarded the structure (Diyana, Jumaidin, Selamat, Alamjuri, et al., 2021).

FTIR Analysis

The FTIR spectra were derived from *Pennisetum purpureum* fibres of varying ages—30, 45, and 60 days. As illustrated in Figure 6, the key peak positions revealed the fibre's composition and the presence of its chemical functional groups. Figure 6 reveals a peak at 3287 cm^{-1} for 30-day aged fibre, showing evidence of O-H bonds in the carboxylic acid group of cellulose components (Ilaiya Perumal & Sarala et al., 2020; Vijay et al., 2020). The following two peaks of *P. purpureum* fibre, found at 2917 and 2850 cm^{-1}, were associated with the C-H stretching vibrations of CH and CH_2 groups, which alluded to the cellulose and hemicellulose presence (Ilyas et al., 2017; Kulandaivel et al., 2020). A peak at 2000 cm^{-1} was linked to the symmetrical stretching of CH_2 in wax, while a smaller peak at 1641 cm^{-1} was associated with the hemicellulose components (Gurukarthik Babu et al.,

2019). The peak at 1517 cm^{-1} captured a minute amount of moisture presence, while the peak at 1033 cm^{-1} was linked to lignin, specifically the C-OH molecules (Manimaran et al., 2020). The presence of absorption bands around 1030–1033 cm^{-1} across all spectra corresponds to the C–O–C stretching vibrations within the pyranose ring, indicative of polysaccharide structures such as cellulose and hemicellulose. This assignment aligns with characteristic infrared bands observed in polysaccharides, where the C–O–C stretching vibrations typically appear in this region (Gieroba et al., 2023). However, the absence of a distinct peak in the 900–950 cm^{-1} range suggests that the C–O.–C glycosidic bond vibrations are not prominently featured in the spectra (Das et al., 2024). Table 6 compares FTIR peak positions and the stretching vibration patterns in *P. purpureum* fibres at the ages of 30, 45, and 60 days.

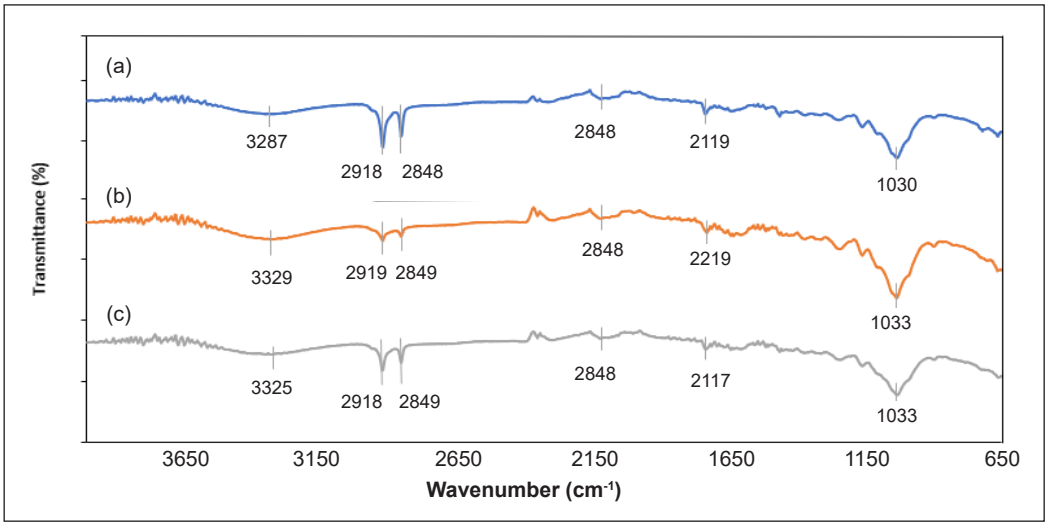


Figure 6. Fourier transform infrared spectroscopy spectrum analyses for *Pennisetum purpureum* fibre for different fibre ages: (a) 30-day; (b) 45-day; and (c) 60-day

Table 6
FTIR peak locations and chemical stretching allocations of Pennisetum purpureum for 30, 45, and 60 days of fibre

Allocations	Fibre age		
	Wavenumber (cm^{-1})		
	30-day	45-day	60-day
OH stretching in α-cellulose	3287	3329	3325
CH stretching in α-cellulose	2918 and 2848	2919 and 2849	2918 and 2849
C=C stretching in wax	2119	2129	2117
C=O stretching in hemicellulose	1731	1728	1734
C-OH vibration in lignin	1030	1033	1033

XRD Analysis

Figure 7 illustrates the X-ray diffractogram for *P. purpureum* fibres at three different ages. The XRD spectra of *P. purpureum* fibres revealed two clear peaks at around 2θ = 16.14° and 22.02° for the 30-day fibres, 16.41° and 22.3° for the 45-day fibres, and 16.39° and 22.3° for the 60-day fibres. The initial peak, located around the 16.14° diffraction angle, corresponded to the components of amorphous presence in *P. purpureum* fibre. The subsequent peak revealed the crystalline component in *P. purpureum*. The elevated CI observed in *P. purpureum* fibre demonstrated superior mechanical capabilities and molecular organisation (Indran et al., 2014). The sharpness of the diffraction peaks reflected the degree of crystallinity, with sharper peaks indicating greater crystallinity in the fibre (Alemdar & Sain, 2008). The CI of *Pennisetum purpureum* fibre was calculated using the Segal empirical technique, as indicated in Equation 4, yielding values of 51.19, 49.40, and 50.56% for fibre ages of 30, 45, and 60 days, respectively. The crystalline sizes were estimated to be 1.68 nm for 30-day age, 1.26 nm for 45-day age, and 1.83 nm for 60-day age. The elevated CI at 30 days, despite a reduced cellulose content relative to the 45-day fibre, can be attributed to the gradual reduction of amorphous constituents such as hemicellulose and lignin. While cellulose is the principal factor in crystallinity, the elimination or degradation of hemicellulose and lignin during plant maturation enhanced the packing density and arrangement of the residual cellulose chains. French (2014) emphasized that less amorphous content promotes the development of more defined crystalline areas in cellulose. Furthermore, lignin enhances the structural rigidity of the fibre matrix, facilitating the alignment of cellulose chains, whilst the semi-crystalline characteristics of hemicellulose assist in integrating cellulose microfibrils into a cohesive structure. These interactions create a stable environment for the formation of crystalline areas in cellulose, notwithstanding the moderate cellulose content. Comparable results

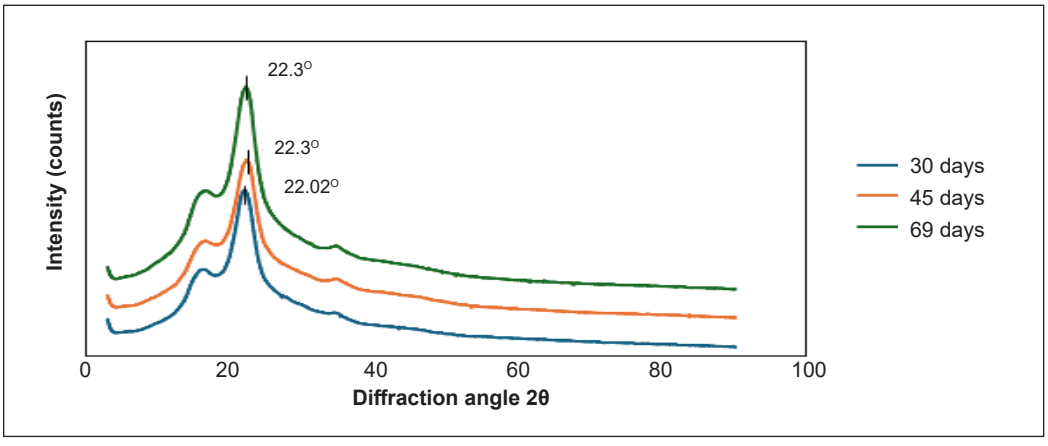


Figure 7. X-ray diffraction trends of *Pennisetum purpureum* fibre for different ages: 30, 45, and 60 days

were documented by Mazumder and Zhang (2023), who observed that the interactions of lignin and hemicellulose with cellulose are crucial for improving crystalline organisation during the initial stages of fibre development.

The CI value of *P. purpureum* fibre surpassed that of most natural fibres, including *H. polyrhizus* (32.76%), *P. amaryllifolius* (37.09%), *C. citratus* (35.2%), *Tridax procumbens* (34.46%), and *J. effusus* L. (33.40%), yet was marginally lower than that of *Pandanus tectorius* (55.10%) and *Cissus quadrangularis* root, as presented in Table 7. Consequently, the crystallinity value was contingent upon the cellulose concentration and the plant species. There exist correlations among fibre stiffness, the degree of crystallinity, and cellulose concentration. Enhancing the crystallinity area augmented the stiffness of the fibres; this can be associated with the correlation of increased fibre crystallinity and cellulose content (Kamaruddin et al., 2021). X-ray diffraction analysis indicated that *P. purpureum* fibre could be a promising material for reinforcing polymer composites.

Table 7
Comparison of the crystallinity of Pennisetum purpureum *fibre and other various fibres*

Fibre	Crystallinity index (%)	Crystallinity size (nm)	References
Pennisetum purpureum - age 30 days	51.19	1.68	Current study
Pennisetum purpureum - age 45 days	49.40	1.26	Current study
Pennisetum purpureum - age 60 days	50.56	1.83	Current study
Hylocereus polyrhizus	32.76		Taharuddin et al. (2023)
*Pandanus amaryllifoliu*s	37.09	4.95	Diyana, Jumaidin, Selamat, Alamjuri, et al. (2021)
Cymbopogon citratus	35.2	-	Kamaruddin et al. (2021)
Tridax procumbens	34.4	25.04	Garette Jebadurai et al. (2019)
Juncus effusus L.	33.40	3.60	Narayanasamy et al. (2020)
Ficus religiosa	42.92	5.18	Chakravarthy K et al. (2020)
Pandanus tectorius	55.10	-	Narayanasamy et al. (2020)
Cissus quadrangularis root	56.60	7.04	Balaji and Nagarajan (2017)

CONCLUSION

This study highlighted that fibres from *P. purpureum* at three different ages (30, 45, and 60 days) were successfully extracted using a water-retting process specific to each age group. These fibres were subsequently analyzed to evaluate their physical, mechanical, thermal, and chemical properties. The density of *P. purpureum* fibre varies with age, influencing its suitability for different applications. The fibres at 30, 45, and 60 days recorded densities of 0.67, 1.03, and 1.23 g/cm^3, respectively, which are lower than many other natural fibres, such as ramie, roselle, and *A. pinnata*. This relatively low density suggests that *P.*

purpureum fibres could be advantageous in producing lightweight composite materials. The fibre of *P. purpureum* showed significant changes in chemical composition depending on the harvesting age. The mechanical properties of *P. purpureum* fibres were not solely dependent on cellulose content but were greatly influenced by the balance and interaction between cellulose and lignin. The optimal mechanical performance at 30 days highlighted the critical role of lignin in providing structural reinforcement (tensile strength: 67.537 MPa, modulus: 3.142 GPa, cellulose: 35.61%, and lignin: 16.47%). The density of the *P. purpureum* fibre increased significantly with age (30, 45, and 60 days), suggesting increased lignification and cell wall compaction over time. Meanwhile, moisture content increased slightly with age, reflecting a potential increase in hydrophilic components like hemicellulose. Thermogravimetric analysis indicated that fibre age influences thermal degradation, with older fibres showing more hemicellulose decomposition and better thermal stability. The 45-day fibre exhibited the lowest moisture loss (4.78%) and the highest char residue (28.54%), indicating better thermal resistance compared to the 30-day and 60-day fibres. The major weight loss due to cellulose and hemicellulose decomposition occurred between 276–390°C, with the 60-day fibre showing a slightly higher peak degradation temperature (390°C), suggesting a more structured cellulose composition. The consistent char residue suggested that the fibres have the potential for applications requiring high thermal resistance. The XRD analysis of *P. purpureum* fibres at different ages revealed distinct crystalline and amorphous regions, with two major peaks appearing consistently around 16° and 22°. The CI values for fibres aged 30, 45, and 60 days were 51.19, 49.40, and 50.56%, respectively, indicating that fibre age influences the crystalline structure. The highest CI value at 30 days suggests that the reduction of amorphous components, such as hemicellulose and lignin, contributed to better molecular organization. Despite variations in cellulose content, the alignment of cellulose chains and interactions with lignin and hemicellulose played a crucial role in enhancing the crystalline structure. This study will be expanded to assess the compatibility of *P. purpureum* fibre with a biodegradable matrix sourced from natural materials.

ACKNOWLEDGEMENTS

The authors would like to thank Universiti Teknikal Malaysia Melaka, Universiti Malaysia Sabah, Universiti Kuala Lumpur, and the Ministry of Higher Education Malaysia for the financial support through the research grant FRGS/1/2023/STG05/UTEM/02/2.

REFERENCES

Abdullah, H. H., Zakaria, S., Anuar, N. I. S., Mohd Salleh, K., & Syeed Jaafar, S. N. (2020). Effect of harvesting time and water retting fiber processing methods on the physico-mechanical properties of kenaf fiber. *BioResources*, *15*(3), 7207–7222. https://doi.org/10.15376/BIORES.15.3.7207-7222

Abera, G., Woldeyes, B., Demash, H. D., & Miyake, G. (2020). The effect of plasticizers on thermoplastic starch films developed from the indigenous Ethiopian tuber crop anchote (*Coccinia abyssinica*) starch. *International Journal of Biological Macromolecules*, *155*, 581–587. https://doi.org/10.1016/j.ijbiomac.2020.03.218

Alemdar, A., & Sain, M. (2008). Isolation and characterization of nanofibers from agricultural residues - Wheat straw and soy hulls. *Bioresource Technology*, *99*(6), 1664–1671. https://doi.org/10.1016/j.biortech.2007.04.029

Arul Marcel Moshi, A., Ravindran, D., Sundara Bharathi, S. R., Padma, S. R., Indran, S., & Divya, D. (2020). Characterization of natural cellulosic fiber extracted from *Grewia damine* flowering plant's stem. *International Journal of Biological Macromolecules*, *164*, 1246–1255. https://doi.org/10.1016/j.ijbiomac.2020.07.225

ASTM International. (2020a). *ASTM C1557-20: Standard test method for tensile strength and Young's modulus of fibers*. https://store.astm.org/c1557-20.html

ASTM International. (2020b). *ASTM D792-20: Standard test methods for density and specific gravity (relative density) of plastics by displacement*. https://store.astm.org/d0792-20.html

Balaji, A. N., & Nagarajan, K. L. (2017). Characterization of alkali treated and untreated new cellulosic fiber from Saharan aloe vera cactus leaves. *Carbohydrate Polymers*, *174*, 200–208. https://doi.org/10.1016/j.carbpol.2017.06.065

Bezazi, A., Belaadi, A., Bourchak, M., Scarpa, F., & Boba, K. (2014). Novel extraction techniques, chemical and mechanical characterisation of *Agave americana* L. natural fibres. *Composites Part B: Engineering*, *66*, 194–203. https://doi.org/10.1016/j.compositesb.2014.05.014

Biradar, A., Arulvel, S., & Kandasamy, J. (2023). Significance of ballistic parameters and nanohybridization in the development of textile-based body armor: A review. *International Journal of Impact Engineering*, *180*, 104700. https://doi.org/10.1016/j.ijimpeng.2023.104700

Boumediri, H., Bezazi, A., Del Pino, G. G., Haddad, A., Scarpa, F., & Dufresne, A. (2019). Extraction and characterization of vascular bundle and fiber strand from date palm rachis as potential bio-reinforcement in composite. *Carbohydrate Polymers*, *222*, 114997. https://doi.org/10.1016/j.carbpol.2019.114997

Campos, A., Sena Neto, A. R., Rodrigues, V. B., Luchesi, B. R., Mattoso, L. H. C., & Marconcini, J. M. (2018). Effect of raw and chemically treated oil palm mesocarp fibers on thermoplastic cassava starch properties. *Industrial Crops and Products*, *124*, 149–154. https://doi.org/10.1016/j.indcrop.2018.07.075

Chakravarthy K, S., Madhu S., Raju, J. S. N., & Shariff Md, J. (2020). Characterization of novel natural cellulosic fiber extracted from the stem of *Cissus vitiginea* plant. *International Journal of Biological Macromolecules*, *161*, 1358–1370. https://doi.org/10.1016/j.ijbiomac.2020.07.230

Chanpla, M., Kullavanijaya, P., Janejadkarn, A., & Chavalparit, O. (2018). Effect of harvesting age and performance evaluation on biogasification from Napier grass in separated stages process. *KSCE Journal of Civil Engineering*, *22*(1), 40–45. https://doi.org/10.1007/s12205-017-1164-y

Chuwongpanich, W., Chittamart, N., Tawornpruek, S., Aramrak, S., & Fujii, K. (2019). *Studies on microbial biomass carbon and nitrogen turnover derived from sugarcane residues incorporated into a sandy loam*. ResearchGate. https://www.researchgate.net/publication/338965599_STUDIES_ON_MICROBIAL_

BIOMASS_CARBON_AND_NITROGEN_TURNOVER_DERIVED_FROM_SUGARCANE_RESIDUES_INCORPORATED_INTO_A_SANDY_LOAM_SOIL

Das, S., Bhati, V., Dewangan, B. P., Gangal, A., Mishra, G. P., Dikshit, H. K., & Pawar, P. A. M. (2024). Combining Fourier-transform infrared spectroscopy and multivariate analysis for chemotyping of cell wall composition in mungbean (*Vigna radiata* (L.) Wizcek). *Plant Methods*, *20*, 135. https://doi.org/10.1186/s13007-024-01260-w

Diyana, Z. N., Jumaidin, R., Selamat, M. Z., Alamjuri, R. H., & Md Yusof, F. A. (2021). Extraction and characterization of natural cellulosic fiber from *Pandanus amaryllifolius* leaves. *Polymers*, *13*(23), 4171. https://doi.org/10.3390/polym13234171

Diyana, Z. N., Jumaidin, R., Selamat, M. Z., Ghazali, I., Julmohammad, N., Huda, N., & Ilyas, R. A. (2021). Physical properties of thermoplastic starch derived from natural resources and its blends: A review. *Polymers*, *13*(9), 1396. https://doi.org/10.3390/polym13091396

Donaldson, L., Nanayakkara, B., & Harrington, J. (2016). Wood growth and development. In B. Thomas, B. G. Murray, & D. J. Murphy (Eds.), *Encyclopedia of applied plant sciences* (2nd ed., Vol. 1, pp. 203–210). Academic Press. https://doi.org/10.1016/B978-0-12-394807-6.00114-3

Dorez, G., Ferry, L., Sonnier, R., Taguet, A., & Lopez-Cuesta, J.-M. (2014). Effect of cellulose, hemicellulose and lignin contents on pyrolysis and combustion of natural fibers. *Journal of Analytical and Applied Pyrolysis*, *107*, 323–331. https://doi.org/10.1016/j.jaap.2014.03.017

Engelbrecht-Wiggans, A. E., & Forster, A. L. (2023). Analysis of strain correction procedures for single fiber tensile testing. *Composites Part A: Applied Science and Manufacturing*, *167*, 107411. https://doi.org/10.1016/j.compositesa.2022.107411

Ferreira, E. A., de Abreu, J. G., Martinez, J. C., Amorim, R. S. S., Neto, A. B., Cabral, C. E. A., dos Santos Braz, T. G., Júnior, C. A. S., & Ferreira, D. P. (2019). Productivity and nutritional value of elephant grass BRS Canará forage. *Semina:Ciencias Agrarias*, *40*(6), 2705–2718. https://doi.org/10.5433/1679-0359.2019v40n6p2705

French, A. D. (2014). Idealized powder diffraction patterns for cellulose polymorphs. *Cellulose*, *21*, 885–896. https://doi.org/10.1007/s10570-013-0030-4

Gallos, A., Paës, G., Allais, F., & Beaugrand, J. (2017). Lignocellulosic fibers: A critical review of the extrusion process for enhancement of the properties of natural fiber composites. *RSC Advances*, *7*(55), 34638–34654. https://doi.org/10.1039/c7ra05240e

Ganapathy, T., Sathiskumar, R., Senthamaraikannan, P., Saravanakumar, S. S., & Khan, A. (2019). Characterization of raw and alkali treated new natural cellulosic fibres extracted from the aerial roots of banyan tree. *International Journal of Biological Macromolecules*, *138*, 573–581. https://doi.org/10.1016/j.ijbiomac.2019.07.136

Garette Jebadurai, S., Edwin Raj, R., Sreenivasan, V. S., & Binoj, J. S. (2019). Comprehensive characterization of natural cellulosic fiber from *Coccinia grandis* stem. *Carbohydrate Polymers*, *207*, 675–683. https://doi.org/10.1016/j.carbpol.2018.12.027

Gieroba, B., Kalisz, G., Krysa, M., Khalavka, M., & Przekora, A. (2023). Application of vibrational spectroscopic techniques in the study of the natural polysaccharides and their cross-linking process. *International Journal of Molecular Sciences*, *24*(3), 2630. https://doi.org/10.3390/ijms24032630

Goda, K., Sreekala, M. S., Gomes, A., Kaji, T., & Ohgi, J. (2006). Improvement of plant based natural fibers for toughening green composites - Effect of load application during mercerization of ramie fibers. *Composites Part A: Applied Science and Manufacturing*, *37*(12), 2213–2220. https://doi.org/10.1016/j.compositesa.2005.12.014

Gurukarthik Babu, B., Prince Winston, D., SenthamaraiKannan, P., Saravanakumar, S. S., & Sanjay, M. R. (2019). Study on characterization and physicochemical properties of new natural fiber from *Phaseolus vulgaris*. *Journal of Natural Fibers*, *16*(7), 1035–1042. https://doi.org/10.1080/15440478.2018.1448318

Huzaifah, M. R. M., Sapuan, S. M., Leman, Z., & Ishak, M. R. (2017). Comparative study on chemical composition, physical, tensile, and thermal properties of sugar palm fiber (*Arenga pinnata*) obtained from different geographical locations. *BioResources*, *12*(4), 9366–9382. https://doi.org/10.15376/biores.12.4.9366-9382

Ilaiya Perumal, C., & Sarala, R. (2020). Characterization of a new natural cellulosic fiber extracted from *Derris scandens* stem. *International Journal of Biological Macromolecules*, *165*(Part B), 2303-2313. https://doi.org/10.1016/j.ijbiomac.2020.10.086

Ilyas, R. A., Sapuan, S. M., Ishak, M. R., & Zainudin, E. S. (2017). Effect of delignification on the physical, thermal, chemical, and structural properties of sugar palm fibre. *BioResources*, *12*(4), 8734–8754. https://doi.org/10.15376/biores.12.4.8734-5754

Ilyas, R. A., Sapuan, S. M., Ishak, M. R., & Zainudin, E. S. (2018). Development and characterization of sugar palm nanocrystalline cellulose reinforced sugar palm starch bionanocomposites. *Carbohydrate Polymers*, *202*, 186–202. https://doi.org/10.1016/j.carbpol.2018.09.002

Imraan, M., Ilyas, R. A., Norfarhana, A. S., Bangar, S. P., Knight, V. F., & Norrrahim, M. N. F. (2023). Sugar palm (*Arenga pinnata*) fibers: New emerging natural fibre and its relevant properties, treatments and potential applications. *Journal of Materials Research and Technology*, *24*, 4551–4572. https://doi.org/10.1016/j.jmrt.2023.04.056

Indran, S., Edwin Raj, R., & Sreenivasan, V. S. (2014). Characterization of new natural cellulosic fiber from *Cissus quadrangularis* root. *Carbohydrate Polymers*, *110*, 423–429. https://doi.org/10.1016/j.carbpol.2014.04.051

Jayaramudu, J., Guduri, B. R., & Varada Rajulu, A. (2010). Characterization of new natural cellulosic fabric *Grewia tilifolia*. *Carbohydrate Polymers*, *79*(4), 847–851. https://doi.org/10.1016/j.carbpol.2009.10.046

Jumaidin, R., Khiruddin, M. A. A., Saidi, Z. A. S., Salit, M. S., & Ilyas, R. A. (2020). Effect of cogon grass fibre on the thermal, mechanical and biodegradation properties of thermoplastic cassava starch biocomposite. *International Journal of Biological Macromolecules*, *146*, 746–755. https://doi.org/10.1016/j.ijbiomac.2019.11.011

Kabir, M. M., Wang, H., Lau, K. T., & Cardona, F. (2012). Chemical treatments on plant-based natural fibre reinforced polymer composites: An overview. *Composites Part B: Engineering*, *43*(7), 2883–2892. https://doi.org/10.1016/j.compositesb.2012.04.053

Kamaruddin, Z. H., Jumaidin, R., Kamaruddin, Z. H., Asyraf, M. R. M., Razman, M. R., & Khan, T. (2023). Effect of *Cymbopogon citratus* fibre on physical and impact properties of thermoplastic cassava starch/palm wax composites. *Polymers*, *15*(10), 2364. https://doi.org/10.3390/polym15102364

Kamaruddin, Z. H., Jumaidin, R., Rushdan, A. I., Selamat, M. Z., & Alamjuri, R. H. (2021). Characterization of natural cellulosic fiber isolated from Malaysian *Cymbopogon citratus* leaves. *BioResources*, *16*(4), 7729–7750. https://doi.org/10.15376/biores.16.4.7729-7750

Karimah, A., Ridho, M. R., Munawar, S. S., Adi, D. S., Ismadi., Damayanti, R., Subiyanto, B., Fatriasari, W., & Fudholi, A. (2021). A review on natural fibers for development of eco-friendly bio-composite: Characteristics, and utilizations. *Journal of Materials Research and Technology*, *13*, 2442–2458. https://doi.org/10.1016/j.jmrt.2021.06.014

Karimah, A., Ridho, M. R., Munawar, S. S., Ismadi., Amin, Y., Damayanti, R., Lubis, M. A. R., Wulandari, A. P., Nurindah., Iswanto, A. H., Fudholi, A., Asrofi, M., Saedah, E., Sari, N. H., Pratama, B. R., Fatriasari, W., Nawawi, D. S., Rangappa, S. M., & Siengchin, S. (2021). A comprehensive review on natural fibers: Technological and socio-economical aspects. *Polymers*, *13*(24), 4280. https://doi.org/10.3390/polym13244280

Khalid, M. Y., Al Rashid, A., Arif, Z. U., Ahmed, W., Arshad, H., & Zaidi, A. A. (2021). Natural fiber reinforced composites: Sustainable materials for emerging applications. *Results in Engineering*, *11*, 100263. https://doi.org/10.1016/j.rineng.2021.100263

Khoo, P. S., Hassan, S. A., Ilyas, R. A., Loganathan, T. G., Adhiguna, R. T., Reby Roy, K. E., & Mubarak Ali, M. (in press). Tensile properties of ramie fibre: Effect of harvesting day and extraction method. *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2023.04.247

Kulandaivel, N., Muralikannan, R., & KalyanaSundaram, S. (2020). Extraction and characterization of novel natural cellulosic fibers from pigeon pea plant. *Journal of Natural Fibers*, *17*(5), 769–779. https://doi.org/10.1080/15440478.2018.1534184

Maache, M., Bezazi, A., Amroune, S., Scarpa, F., & Dufresne, A. (2017). Characterization of a novel natural cellulosic fiber from *Juncus effusus* L. *Carbohydrate Polymers*, *171*, 163–172. https://doi.org/10.1016/j.carbpol.2017.04.096

Maideliza, T., Mayerni, R., & Rezki, D. (2017). Comparative study of length and growth rate of ramie (*Boehmeria nivea* L. Gaud.) bast fiber of Indonesian clones. *International Journal on Advanced Science, Engineering and Information Technology*, *7*(6), 2273–2278.

Manimaran, P., Pillai, G. P., Vignesh, V., & Prithiviraj, M. (2020). Characterization of natural cellulosic fibers from Nendran Banana Peduncle plants. *International Journal of Biological Macromolecules*, *162*, 1807–1815. https://doi.org/10.1016/j.ijbiomac.2020.08.111

Mazumder, S., & Zhang, N. (2023). Cellulose–hemicellulose–lignin interaction in the secondary cell wall of coconut endocarp. *Biomimetics*, *8*(2), 188. https://doi.org/10.3390/biomimetics8020188

Mohammed, M., Mohamad Jawad, A. J., Mohammed, A. M., Oleiwi, J. K., Adam, T., Osman, A. F., Dahham, O. S., Betar, B. O., Gopinath, S. C. B., & Jaafar, M. (2023). Challenges and advancement in water absorption of natural fiber-reinforced polymer composites. *Polymer Testing*, *124*, 108083. https://doi.org/10.1016/j.polymertesting.2023.108083

Narayanasamy, P., Balasundar, P., Senthil, S., Sanjay, M. R., Siengchin, S., Khan, A., & Asiri, A. M. (2020). Characterization of a novel natural cellulosic fiber from *Calotropis gigantea* fruit bunch for ecofriendly

polymer composites. *International Journal of Biological Macromolecules*, *150*, 793–801. https://doi.org/10.1016/j.ijbiomac.2020.02.134

Oladele, I. O., Michael, O. S., Adediran, A. A., Balogun, O. P., & Ajagbe, F. O. (2020). Acetylation treatment for the batch processing of natural fibers: Effects on constituents, tensile properties and surface morphology of selected plant stem fibers. *Fibers*, *8*(12), 73. https://doi.org/10.3390/fib8120073

Onjai-uea, N., Paengkoum, S., Taethaisong, N., Thongpea, S., Sinpru, B., Surakhunthod, J., Meethip, W., Purba, R. A. P., & Paengkoum, P. (2023). Effect of cultivar, plant spacing and harvesting age on yield, characteristics, chemical composition, and anthocyanin composition of purple Napier grass. *Animals*, *13*(1), 10. https://doi.org/10.3390/ani13010010

Purba, R. A. P., & Paengkoum, P. (2019). Bioanalytical HPLC method of *Piper betle* L. for quantifying phenolic compound, water-soluble vitamin, and essential oil in five different solvent extracts. *Journal of Applied Pharmaceutical Science*, *9*(5), 33–39. https://doi.org/10.7324/JAPS.2019.90504

Radakisnin, R., Abdul Majid, M. S., Mohd Jamir, M. R., Jawaid, M., Sultan, M. T. H., & Mat Tahir, M. F. (2020). Structural, morphological and thermal properties of cellulose nanofibers from napier fiber (*Pennisetum purpureum*). *Materials*, *13*(18), 4125. https://doi.org/10.3390/ma13184125

Razali, N., Salit, M. S., Jawaid, M., Ishak, M. R., & Lazim, Y. (2015). A study on chemical composition, physical, tensile, morphological, and thermal properties of roselle fibre: Effect of fibre maturity. *BioResources*, *10*(1), 1803–1824. https://doi.org/10.15376/biores.10.1.1803-1824

Reddy, N., & Yang, Y. (2005). Biofibers from agricultural byproducts for industrial applications. *Trends in Biotechnology*, *23*(1), 22–27. https://doi.org/10.1016/j.tibtech.2004.11.002

Sheltami, R. M., Abdullah, I., Ahmad, I., Dufresne, A., & Kargarzadeh, H. (2012). Extraction of cellulose nanocrystals from mengkuang leaves (*Pandanus tectorius*). *Carbohydrate Polymers*, *88*(2), 772–779. https://doi.org/10.1016/j.carbpol.2012.01.062

Sultana, R. S., & Rahman, M. M. (2014). An overview of microfibril angle in fiber of tension wood. *European Journal of Biophysics*, *2*(2), 7-12. https://doi.org/10.11648/j.ejb.20140202.11

Sunny, G., & Rajan, T. P. (2022). Review on areca nut fiber and its implementation in sustainable products development. *Journal of Natural Fibers*, *19*(12), 4747–4760. https://doi.org/10.1080/15440478.2020.1870623

Suryanto, H., Marsyahyo, E., Irawan, Y. S., & Soenoko, R. (2014). Morphology, structure, and mechanical properties of natural cellulose fiber from Mendong grass (*Fimbristylis globulosa*). *Journal of Natural Fibers*, *11*(4), 333–351. https://doi.org/10.1080/15440478.2013.879087

Taharuddin, N. H., Jumaidin, R., Mansor, M. R., Md Yusof, F. A., & Alamjuri, R. H. (2023). Characterization of potential cellulose from *Hylocereus polyrhizus* (dragon fruit) peel: A study on physicochemical and thermal properties. *Journal of Renewable Materials*, *11*(1), 131–145. https://doi.org/10.32604/jrm.2022.021528

Tarique, J., Sapuan, S. M., Khalina, A., Sherwani, S. F. K., Yusuf, J., & Ilyas, R. A. (2021). Recent developments in sustainable arrowroot (*Maranta arundinacea* Linn) starch biopolymers, fibres, biopolymer composites and their potential industrial applications: A review. *Journal of Materials Research and Technology*, *13*, 1191–1219. https://doi.org/10.1016/j.jmrt.2021.05.047

Todkar, S. S., & Patil, S. A. (2019). Review on mechanical properties evaluation of pineapple leaf fibre (PALF) reinforced polymer composites. *Composites Part B: Engineering*, *174*, 106927. https://doi.org/10.1016/j.compositesb.2019.106927

Vijay, R., Singaravelu, D. L., Vinod, A., Paul Raj, I. D. F., Sanjay, M. R., & Siengchin, S. (2020). Characterization of novel natural fiber from *Saccharum bengalense* grass (Sarkanda). *Journal of Natural Fibers*, *17*(12), 1739–1747. https://doi.org/10.1080/15440478.2019.1598914

Wahid, M. K., Jumaidin, R., Mohd Suan, M. S., & Md Yusof, F. A. (2023). Characterisation of *Axonopus compressus* fibre through chemical composition, thermogravimetric behaviour, and moisture content evaluation. *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences*, *110*(1), 172–181. https://doi.org/10.37934/arfmts.110.1.172181

Yuan, J., Liu, G., Liu, P., & Huang, R. (2024). Comprehensive assessment of elephant grass (*Pennisetum purpureum*) stalks at different growth stages as raw materials for nanocellulose production. *Tropical Plants*, *3*, e013. https://doi.org/10.48130/tp-0024-0013

Zaini, N., Tilova, A. M., Umami, N., Hanim, C., Astuti, A., & Suwignyo, B. (2021). Effect of harvesting age of chicory (*Cichorium intybus*) on the pattern of planting intercropping dwarf elephant grass in the second regrowth on production and quality. In *IOP Conference Series: Earth and Environmental Science* (Vol. 788, No. 1, p. 012173). IOP Publishing. https://doi.org/10.1088/1755-1315/788/1/012173

Zhang, B., Guo, Y., Liu, X., Chen, H., Yang, S., & Wang, Y. (2020). Mechanical properties of the fiber cell wall in *Bambusa pervariabilis* bamboo and analyses of their influencing factors. *BioResources*, *15*(3), 5316–5327. https://doi.org/10.15376/biores.15.3.5316-5327

Zhang, W., Zhang, S., Lu, X., Li, C., Liu, X., Dong, G., & Xia, T. (2020). Tissue-specific transcriptome analysis reveals lignocellulose synthesis regulation in elephant grass (*Pennisetum purpureum* Schum). *BMC Plant Biology*, *20*, 528. https://doi.org/10.1186/s12870-020-02735-3

Zor, M., Şen, F., Yazıcı, H., & Candan, Z. (2023). Thermal, mechanical and morphological properties of cellulose/lignin nanocomposites. *Forests*, *14*(9), 1715. https://doi.org/10.3390/f14091715

Pertanika J. Sci. & Technol. 33 (6): 2491 - 2520 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Harnessing Co-Culture of Hydrolase–Producing *Bacillus* spp. to Boost Kitchen Waste Biodegradation

Nurhidayah Ramlee[1], Muhamad Aidilfitri Mohamad Roslan[2], Helmi Wasoh[1] and Nor Aini Abdul Rahman[1]*

[1]Department of Bioprocess Technology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[2]Department of Crop Science, Faculty of Agricultural and Forestry Sciences, Universiti Putra Malaysia, Bintulu Sarawak Campus, Nyabau Road, 97008 Bintulu, Sarawak, Malaysia

ABSTRACT

In addressing the urgent challenge of organic waste management, our study investigates the use of hydrolase-producing bacteria (HPB) co-cultures as a targeted approach to enhance the biodegradation of kitchen waste (KW). This study aimed to isolate HPB with high enzyme activity to facilitate more efficient kitchen waste breakdown. The screening of 26 bacterial strains from fermented vegetables, agricultural soils, and in-house glycerol stock collections revealed *Bacillus licheniformis* 2D55 and *Bacillus xiamenensis* Y7 as standout candidates, exhibiting multiple hydrolase activities, including amylase, protease, lipase, and cellulase. In co-culture fermentation, these strains demonstrated superior hydrolase activities, microbial count, percentage of biodegraded total solids (TS), and gross degradation rate (GDR) during solid-state fermentation (SSF) compared to monocultures and controls. The co-culture treatments achieved 54.73% and 65% biodegraded TS in sterile and non-sterile SSF, respectively. Additionally, a 25.39% GDR in sterile SSF and 41.36% in non-sterile SSF were observed after the 14-day fermentation process. In the composting experiment, the inoculated compost exhibited a higher biodegraded TS (63.02%) compared to the control (29.80%). These findings underscore the potential of HPB co-cultures as an effective strategy to improve KW biodegradation, contributing to sustainable organic waste management solutions.

Keywords: Biodegradation, composting, hydrolase-producing bacteria, kitchen waste, solid-state fermentation

ARTICLE INFO

Article history:
Received: 08 June 2024
Accepted: 31 January 2025
Published: 30 September 2025

DOI: https://doi.org/10.47836/pjst.33.6.05

E-mail addresses:
nurhidayahramlee08@gmail.com (Nurhidayah Ramlee)
aidilfitri@upm.edu.my (Muhamad AidilfItri Mohamad Roslan)
helmi_wmi@upm.edu.my (Helmi Wasoh)
nor_aini@upm.edu.my (Nor Aini Abdul Rahman)
* Corresponding author

INTRODUCTION

In developing countries like Malaysia, the disposal of household waste, a substantial component of municipal solid waste (MSW), predominantly occurs in landfills,

e-ISSN: 2231-8526
© Universiti Putra Malaysia Press

many of which are anticipated to reach capacity within a decade (Abubakar et al., 2022). The complexity of managing MSW in Malaysia is exacerbated by escalating greenhouse gas (GHG) emissions from organic waste decomposition in landfills and the heightened demand for material and energy consumption to sustain the nation's growing population (Yong et al., 2019). Therefore, there is a pressing need for quick and practical ways to implement a more thorough and sustainable waste management strategy (Zainu & Songip, 2017).

Solid-state fermentation (SSF) and composting are widely recognized as effective, environmentally friendly approaches for treating organic waste, including kitchen waste (KW). SSF is particularly well-suited for KW due to its operational advantages, such as high microbial density, efficient enzyme production, and the ability to handle substrates with low moisture content, which aligns well with the high organic load in KW (Yafetto, 2022). SSF involves microbial growth on solid substrates, often producing a range of extracellular enzymes, such as hydrolases, that degrade complex organic matter into simpler compounds (Sarkar et al., 2016). Composting, on the other hand, has long been established as a natural, aerobic process for converting organic waste into stable humus-like substances through microbial activity (Storino et al., 2016). The composting process involves microbial hydrolysis and mineralization of organic matter, releasing fewer pollutants compared to other treatment options, and it has been shown to effectively reduce pathogenic microorganisms, promoting sanitary waste disposal (Wei et al., 2017). Given the challenges in landfill capacity and environmental concerns, composting also represents a sustainable alternative that contributes to a circular economy by recycling organic nutrients back into the soil.

The application of HPB in waste treatment has gained increasing attention, particularly in enhancing the biodegradation efficiency of composting and SSF processes. Previous studies have shown that HPB, including strains of *Bacillus* and other genera, contribute significantly to the degradation of complex organic molecules such as starch, proteins, lipids, and cellulose in organic waste (Cerda et al., 2017). For instance, Singh et al. (2021) demonstrated that *Bacillus* spp. with robust amylase and protease activities effectively degraded food waste in a composting setup, resulting in reduced total solids and improved compost quality. This suggests that targeted microbial inoculation could accelerate the composting process and reduce greenhouse gas emissions associated with the decomposition of organic waste. Besides, Drissi Kaitouni et al. (2020) also reported that HPB in SSF processes not only improves enzymatic activity but also enhances the overall microbial biomass, leading to more efficient substrate degradation. However, despite the apparent advantages, only a few studies have been conducted to examine the use of HPB to target KW. The specific data on the KW and the combined effect of HPB with multiple hydrolase activities in both SSF and composting remain sparse (Cerda et al., 2017).

Therefore, the main objective of this research is to isolate and evaluate HPB as an inoculant to improve the biodegradation of KW, which is mainly comprised of rice and vegetables. Rice and vegetables are staple foods in many diets worldwide, so they are often present in significant amounts in KW. Both rice and vegetables present diverse organic components, including carbohydrates, proteins, and fibers. This diversity allows for a comprehensive assessment of the effectiveness of HPB in degrading different substrates commonly found in kitchen waste (Wan et al., 2020). The selected HPB were then tested in sterile and non-sterile SSF. This study evaluates both sterile and non-sterile fermentation conditions to understand the biodegradation potential of monoculture versus co-culture HPB in controlled and natural microbial environments. The sterile fermentation provides a controlled baseline that eliminates external microbial influences, while the non-sterile fermentation introduces natural microbial communities that better mimic the actual environment where biodegradation would occur, such as in soil or waste treatment facilities. This allows us to assess the practical applicability of the co-culture in broader waste management contexts. Lastly, the research evaluated the effects of HPB on the composting of KW.

MATERIALS AND METHODS

Sample Collection

Fermented vegetables, specifically spinach, were obtained from a bioreactor (Biostat® B, Germany) set up in the laboratory experiment for undergraduate students at the Universiti Putra Malaysia (UPM). Agricultural soil, collected from the same research center at a depth of 5–10 cm, was sealed in polyethylene bags. Both samples were stored in 50 ml sterile centrifuge tubes in the chiller (4°C) until use. *Bacillus licheniformis* 2D55 was revived from the glycerol stock solution preserved at -80°C (Kazeem et al., 2016).

Isolation and Screening of Hydrolase-Producing Microorganisms

Isolation was carried out using the serial dilution technique. The samples were transferred to de Man, Rogosa, and Sharpe (MRS) agar (Oxoid, England), nutrient agar (Oxoid, England), yeast extract peptone dextrose (YEPD) agar (BD Biosciences, Canada), and incubated for 24 hr at 37°C. Morphologically distinct isolates were randomly selected and streaked onto freshly prepared agar plates. Each isolate underwent primary screening for hydrolytic enzyme production (amylase, protease, lipase, and cellulase). Starch, casein, tributyrin, and carboxymethylcellulose (CMC) were used in hydrolytic plate assay tests for amylolytic, proteolytic, lipolytic, and cellulolytic activities, respectively. Following incubation for 48 hr at 37°C, specific staining solutions were applied to the culture media. The presence of the corresponding enzymes was indicated by clear zones surrounding the colonies. For amylolytic activity, iodine solution was poured over the isolates after

they were inoculated onto starch agar media (HiMedia, India) (20 g/L soluble starch, 5 g/L peptone, 5 g/L sodium chloride [NaCl], and 20 g/L agar; pH 7) (Simair et al., 2017). Isolates were grown on skim milk agar (Oxoid, England) (28 g/L skim milk powder, 2.5 g/L yeast extract, 1 g/L dextrose, 5 g/L casein, 20 g/L agar; pH 7) to detect proteolytic activity (Tallapragada et al., 2018).

Similarly, isolates were inoculated on spirit blue agar (HiMedia, India) (10 g/L tryptone, 5.0 g/L yeast extract, 0.15 g/L spirit blue, 20 g/L agar; pH 7) mixed with 30 ml of a lipoidal solution (Merck, Germany) (1 ml of polysorbate 80, 100 ml of olive, and 400 ml of warm water at 60°C) to assess lipolytic activity (Privé et al., 2015). For cellulolytic activity, CMC-agar plates (HiMedia, India) (5 g/L CMC, 1 g/L peptone, 2 g/L monopotassium phosphate [KH_2PO_4], 1.4 g/L ammonium sulfate [$(NH_4)_2SO_4$], 0.3 g/L urea, 0.3 g/L magnesium sulfate heptahydrate [$MgSO_4.7H_2O$], 0.3 g/L calcium chloride [$CaCl_2$], 0.0016 g/L manganese (II) sulfate [$MnSO_4$], 20 g/L agar; pH 7) were used to inoculate the isolates. After incubation for 48 hr at 37°C, the plates were flooded for 5 min with Congo red solution (HiMedia, India; 2 g/L), followed by a 15-min wash with 1 M NaCl (Sethi et al., 2013). The existence of a clear zone enclosing the bacterial colony in all assays indicated positive hydrolytic activity (+), whereas the absence of such a zone indicated negative hydrolytic activity (–). The hydrolysis index (HI) was calculated by interpreting the ratio of the hydrolysis zone diameter to the colony diameter (Roslan et al., 2021). All experiments were conducted in triplicate to ensure the reproducibility of results, and the data presented reflect the mean values with standard deviations.

Quantitative Hydrolase Activity Assay

Isolates were introduced into a broth medium containing 1% (w/v) soluble starch (Merck, Germany), 1% (w/v) casein (Merck, Germany), p-nitrophenylpalmitate (p-NPP) (Thermo Scientific, USA), and 1% (w/v) CMC (HiMedia, India) as substrates for evaluating amylolytic, proteolytic, lipolytic, and cellulolytic activities, respectively. The broth underwent incubation for 24 hr at 37°C with shaking at 150 rpm. Subsequently, the culture was centrifuged for 10 min at 11,200 × *g*, and the supernatant was collected to serve as a crude enzyme. Hydrolytic enzyme activities were determined following standard methods. Hydrolase activity assays were performed in triplicate to ensure reproducibility. Identical conditions, such as temperature, pH, and substrate concentrations, were maintained across replicates. Data are presented as mean values with standard deviations to reflect observed variations. Cellulase and amylase activity were assessed using the 3,5-dinitrosalicylic acid (DNS) method (Miller, 1959). Using casein as a substrate, protease activity was determined (Tsuchida et al., 1986); whereas lipase activity was assessed by hydrolyzing p-NPP (Winkler & Stuckmann, 1979). Strains exhibiting more than one enzyme and higher enzymatic activities compared to others were selected and identified.

Characterization and Identification of Hydrolase-Producing Microorganisms

Observation of Gram's reaction and morphological characteristics of the selected strains was conducted using light microscopy (Olympus CH, Japan). The chosen bacteria were cultured overnight at 37°C in nutrient agar media to obtain a pure culture. DNA extraction was conducted using the GF-1 Bacterial DNA extraction kit (QIAGEN, Germany), and the amplification of the 16S rDNA gene fragment was achieved through polymerase chain reaction (PCR) using universal primers 1492R and 27F (Yadav et al., 2009). Using the QIA Quick PCR Purification kit (QIAGEN, Germany), the PCR product was purified. Sequencing of the PCR product was performed by the sequencing service center. To determine the species, the 16S rRNA gene sequences were run via the Basic Local Alignment Search Tool (BLAST). The construction of the phylogenetic tree was accomplished using MEGA X Alignment Explorer (version 10.2.6) (Kumar et al., 2018).

Biocompatibility Assay

The pour-plated method was used to evaluate the biocompatibility of the selected strains (Grossart et al., 2004). In this procedure, a 50 µl suspension of the target bacterial strain (1×10^8 CFU/ml) was combined with molten nutrient 1% agar (Oxoid, England) (2.5 ml). Once transferred onto a nutrient agar plate, the cell suspension agar solidified over a 10-min period. A 10 µl aliquot of the test strain (1×10^8 CFU/ml) was spotted onto the agar surface and incubated for 3 days at 37°C. The assessment of biocompatibility involved measuring the size of the inhibition zone formed on the plates post-incubation. Inhibition zones were monitored daily, and strains were tested in triplicate against one another. When the inhibition zone surrounding the spotted colony measured more than 4 mm in diameter, inhibitory activity was recorded.

Enzymatic Degree of Synergy (DS)

One ml aliquot of the starter culture, with the inoculum adjusted to 1×10^8 CFU/ml and grown in nutrient media, was utilized to inoculate specified production media for each enzyme. Prior to inoculation into the test media, individual inoculum cultures (0.5 ml each) were combined in an equal ratio of 1:1 (v/v) to create the co-culture. The cell-free culture supernatant, grown separately in those selective media, was shaken at 150 rpm for 24 hr at 37°C before the hydrolytic enzyme activity was determined. To obtain the cell-free culture supernatant, the culture was centrifuged for 10 min at 11,200 x *g*. By dividing the observed activity of the co-culture enzyme by the total of all the monocultural enzymes in the same media, the enzymatic DS is calculated (Van Dyk et al., 2013). To assess the effect of pH, each pure bacterial culture and a co-culture of the selected strain were incubated at different pH values ranging from 5.0 to 9.0, using standard assay methods as previously described. Each pure bacterial culture (pH 7) and co-culture of selected strains were incubated for 1

day at temperatures ranging from 30 to 50°C, and the effect of temperature was evaluated. The enzymatic activities were also measured using similar methods.

Solid-State Fermentation

The SSF experiment was conducted in 250 ml Erlenmeyer flasks, each containing 100 g of KW, consisting of rice and vegetable waste collected from a restaurant near Sri Serdang, Selangor, Malaysia. The KW underwent washing, drying, and coarse grinding. The mixture's main characteristics are outlined in Table 1. Two types of pretreatments were employed: sterile KW samples (autoclaved for 15 min at 121°C) and non-sterile KW samples. Rice and vegetable wastes (50 g each) were weighed, mixed, and supplemented with sterilized water to maintain a 60-65% moisture content (Kaewlaoyoong et al., 2020). The mixture was then placed into a sterile conical flask. To prepare the inoculum for SSF, the chosen strains were grown separately in nutrient broth overnight. Subsequently, 1 ml (~1×10^8 CFU/ml) of the inoculum was added to each test flask and incubated statically for 14 days at 25 ± 2°C. For co-culture inoculation, the inoculum was prepared by combining equal volumes of different monocultures (1:1 ratio, v/v) before introducing it into the test flasks. Control flasks received 1 ml of sterile broth each.

Moisture and ash content were determined using the drying method (Kutsanedzie et al., 2012), while pH was monitored daily with a digital pH meter (Thermo Scientific Eutech Instruments, USA). Microbial population during KW degradation was assessed by mixing 9 ml of distilled water containing 0.85% (w/v) NaCl (Merck, Germany) with 1 g of the sample (Varma et al., 2017). Enumeration of the microbial population was conducted on nutrient agar using the standard plate count method (Sanders, 2012). The percentage of KW degradation based on total solids (TS) over the 14-day incubation period was determined following the standard procedure (Leege, 1998). Enzymatic activities (lipase, cellulase, protease, and amylase) were assayed using KW extracts in various solutions of buffer (Leow et al., 2018), and the hydrolase activities were evaluated as per standard procedures. The gross degradation rate (GDR) of KW during the 14-day digestion period was calculated using daily gross weight variations from the initial reading, following the method outlined by T. Li et al. (2014).

Table 1
Properties of the selected kitchen waste components

Parameter	Rice	Vegetables	Rice and vegetables
Moisture (%)	3.62 ± 0.12	15.54 ± 0.24	37.29 ± 0.30
Total solids (%)	97.30 ± 1.16	14.29 ± 2.20	76.31 ± 2.37
pH	6.34 ± 0.02	6.71 ± 0.03	6.68 ± 0.02

Note. Data represent mean ± SD; n = 3

Preparation, Monitoring, and Evaluation of Composting Process

A dual-chamber tumbling composter (Miracle-Gro®, United Kingdom) served as the experimental apparatus for this study. Composting materials, including vegetable waste, goat manure, and rice husk, were sourced from Pasar Borong Selangor, Fakulti Pertanian, UPM, and Kilang Beras Bernas Sekinchan, respectively. Each chamber was filled with 25 kg of feedstock (wet weight), comprising 50% vegetable waste, 25% goat manure, and 25% rice husk, establishing an initial carbon-to-nitrogen (C/N) ratio of 25-30:1. An amount of 1.5 L inoculum for the composting process was prepared by first growing the HPB in 250 ml of nutrient broth at 37°C with shaking for 24 hr. The culture was then diluted by taking 15 ml of the initial culture (~1×10^9 CFU/ml) and mixing it with 235 ml of sterile distilled water. This dilution was further scaled up by adding 1.25 L of sterile distilled water to achieve a final volume of 1.5 L.

The inoculum was mixed thoroughly using a magnetic stirrer to ensure uniform distribution of bacteria throughout the solution. To verify the final bacterial concentration, colony-forming unit (CFU) counts were performed by serial dilution and plating onto agar plates to confirm the final bacterial concentration, which was ~1×10^8 CFU/ml in the inoculum. Optical density (OD) measurements at 600 nm were also taken from aliquots of the inoculum to monitor the growth trend and ensure uniform bacterial distribution. In one chamber, 1.5 L of HPB was added, while in the control chamber, 1.5 L of distilled water containing 15 ml of sterile nutrient broth (prepared in the same manner as for the inoculated treatment) was used to achieve an initial moisture content of 40-60% (Wu et al., 2017). Throughout the two-month composting process, various parameters were monitored, including pH, temperature, odor, color, C/N ratio, hydrolase activities, bacterial population, and the percentage of degradation rate based on TS. Moisture, pH, temperature, hydrolase activities, bacterial population, and TS degradation rate were assessed using methods outlined in the Solid-state Fermentation section. For C/N ratio analysis, samples were dried at 70°C, ground, and analyzed using a carbon, nitrogen, and sulfur (CNS) elemental analyzer (TruMac CNS Macro Analyzer, USA) (Fan et al., 2018).

To evaluate phytotoxicity, a germination test using radish seeds was conducted on the final compost samples, both with and without HPB, over 8 weeks (Siles-Castellano et al., 2020). For each sample, 15 ml of distilled water was mixed with 1.5 g of compost (dry basis) for one hour at room temperature. The mixture was then centrifuged at 2,800 x *g* for 20 min, and after the supernatant was filtered, 5 ml of the extract was added to a container containing 20 radish seeds on tissue paper. Each sample was incubated in the dark for 72 hr at room temperature with two duplicates. As a control test, 20 radish seeds were placed in a container with 5 ml of distilled water. The germination index (GI) was determined based on the number of germinated seeds and the length of their roots (Wang et al., 2022).

Statistical Analysis

Statistical analyses were conducted using two-tailed, unpaired t-tests at a confidence level of $P < 0.05$ and $P < 0.01$ to ascertain the relevance of variations in HPB's HI on screening plates and in the level of HPB synergism. Changes in hydrolase activity across treatments for HPB, TS, and GDR were assessed using one-way analysis of variance (ANOVA). Tukey's test was performed on the calculated mean at $P < 0.05$ and $P < 0.01$. The correlation coefficients between HI and hydrolase activity, hydrolase activity and pH, and hydrolase activity and temperature were measured using Pearson's correlation, all at $P < 0.05$ and $P < 0.01$ confidence ranges. By using SPSS software Package Version 25.0 (SPSS Inc., USA), all statistical analyses were performed.

RESULTS

Isolation and Screening of Hydrolase-Producing Microorganisms

A total of 26 bacterial cultures, comprising 10 isolates from fermented vegetables, 15 isolates from soil samples, and one isolate from a glycerol stock solution, were selected based on diverse colony morphology for further analysis. Screening on plates containing starch, casein, tributyrin, and CMC agar revealed that 24 isolates exhibited significant hydrolase activity (Table 2). Out of these, 16 isolates could produce protease, 8 isolates demonstrated lipase activity, 7 isolates showed cellulase activity, and 6 isolates exhibited amylase activity. Some strains displayed more than one hydrolytic activity, with only strains 2D55 and Y7 showcasing more than two hydrolytic activities. Strain 2D55 displayed all four hydrolytic activities, while strain Y7 produced amylase, protease, and lipase. On average, the HI of 2D55 was consistently higher than that of Y7 in the amylase and lipase agar assays ($P < 0.05$). However, for the protease assay, the HI of 2D55 and Y7 showed no significant difference. Among all hydrolases, 2D55 exhibited the highest HI for cellulase (5.24), while Y7 demonstrated the highest HI for lipase (3.17).

Table 2
Qualitative screening of hydrolytic enzymes produced by the isolates

Source	Isolate	Hydrolysis index			
		Amylase	Protease	Lipase	Cellulase
Fermented vegetables	N1	1.07 ± 0.02a	nd	nd	nd
	N2	nd	1.07 ± 0.02a	1.07 ± 0.02a	nd
	N3	nd	nd	nd	1.07 ± 0.02a
	N4	nd	nd	1.12 ± 0.07a	nd
	N5	1.12 ± 0.07a	1.12 ± 0.07a	nd	nd
	N6	nd	1.08 ± 0.03a	nd	nd
	N7	1.08 ± 0.03a	nd	nd	1.12 ± 0.07a
	N8	nd	nd	nd	3.17 ± 0.11c

Table 2 *(continue)*

Source	Isolate	Hydrolysis index			
		Amylase	Protease	Lipase	Cellulase
	N9	1.21 ± 0.11a	nd	nd	nd
	N10	nd	1.21 ± 0.11ab	nd	nd
Soil sample	Y1	nd	1.38 ± 0.07ab	nd	nd
	Y2	nd	nd	nd	1.21 ± 0.11a
	Y3	nd	nd	nd	nd
	Y4	nd	2.04 ± 0.26c	nd	nd
	Y5	nd	nd	nd	1.38 ± 0.07a
	Y6	nd	nd	nd	nd
	Y7	1.38 ± 0.07a	2.02 ± 0.03c	3.17 ± 0.11c	nd
	Y8	nd	1.53 ± 0.03b	nd	2.04 ± 0.26b
	Y9	nd	1.36 ± 0.14ab	1.21 ± 0.11a	nd
	Y10	nd	1.21 ± 0.09ab	1.38 ± 0.07a	nd
	Y11	nd	1.40 ± 0.21ab	nd	nd
	Y12	nd	1.23 ± 0.08ab	2.04 ± 0.26b	nd
	Y13	nd	1.17 ± 0.04a	nd	nd
	Y14	nd	1.06 ± 0.01a	2.02 ± 0.03b	nd
	Y15	nd	1.11 ± 0.08a	nd	nd
Glycerol stock collection	2D55	2.04 ± 0.26b	2.20 ± 0.02c	4.12 ± 0.24d	5.24 ± 0.24d

Note. Data represent mean ± SD. According to Tukey's test, values in the same column that are followed by different lowercase letters indicate significant differences at $P < 0.05$; nd = Not detected; n = 3

Quantitative Assay of Microbial Hydrolase Activities

In the liquid-based assays (Table 3), strain 2D55 exhibited the highest amylase potential with an enzyme activity and specific activity of 211.06 U/ml and 222.17 U/mg, respectively. The highest protease activity was observed in strain Y1, but no significant difference was found compared to strains 2D55 and Y7. Both strain 2D55 and Y7 showed equally high lipase activity ($P > 0.05$), recording enzyme activities of 303.89 and 313.39 U/ml, respectively. In terms of cellulase activity, strain 2D55 demonstrated the highest enzyme activity, almost double the value of strain N3, at 169.04 U/ml. Additionally, the HI of amylase ($r = 0.887$, $P < 0.05$), protease ($r = 0.578$, $P < 0.05$), lipase ($r = 0.595$, $P < 0.05$), and cellulase ($r = 0.699$, $P < 0.05$) showed significant correlations with hydrolase activity in liquid media. Therefore, strains 2D55 and Y7 were chosen for additional analysis based on the overall outcomes.

Bacterial Characterization and Identification

By conducting biochemical analysis and sequencing the 16S rRNA gene of each strain, the phenotypic and genotypic characteristics of the selected strains were determined (Table 4). The results revealed that both HPBs were capable of thriving in high-temperature

Table 3
Quantitative screening of hydrolytic enzyme-producing isolates

Strain ID		Enzyme activity (U/ml)	Total protein (mg/ml)	Specific activity (U/mg)
Amylase	N1	23.01 ± 7.26a	0.65 ± 0.04b	35.40a
	N5	87.07 ± 3.16c	0.55 ± 0.02a	159.27c
	N7	61.58 ± 5.47b	0.84 ± 0.04c	73.02b
	N9	62.27 ± 3.16b	0.64 ± 0.02b	97.81b
	Y7	149.07 ± 7.82d	0.87 ± 0.02cd	171.34c
	2D55	211.06 ± 6.64e	0.95 ± 0.04d	222.17d
Protease	N2	17.09 ± 0.60a	0.32 ± 0.02a	52.87c
	N5	25.10 ± 6.80a	0.47 ± 0.03bc	53.79cd
	N6	31.37 ± 5.75a	0.60 ± 0.02cd	51.99c
	N10	90.90 ± 3.36b	0.96 ± 0.02g	94.69g
	Y1	118.76 ± 2.76c	1.24 ± 0.02i	95.52g
	Y4	117.71 ± 8.92c	1.35 ± 0.05j	87.41f
	Y7	105.53 ± 6.38bc	0.85 ± 0.03ef	123.67h
	Y8	103.90 ± 14.62bc	1.40 ± 0.02j	73.46e
	Y9	112.84 ± 3.95c	1.53 ± 0.03k	73.91e
	Y10	117.02 ± 18.40c	1.17 ± 0.02h	100.30g
	Y11	23.71 ± 1.04a	1.19 ± 0.04i	19.87a
	Y12	25.45 ± 5.26a	0.54 ± 0.02c	47.42c
	Y13	27.89 ± 2.76a	0.46 ± 0.03b	60.19d
	Y14	39.73 ± 9.48a	0.46 ± 0.01b	86.36f
	Y15	27.89 ± 2.76a	0.78 ± 0.02e	35.60b
	2D55	100.65 ± 3.19bc	0.86 ± 0.02f	117.49h
Lipase	N2	104.46 ± 37.99a	1.36 ± 0.04de	76.62a
	N4	284.90 ± 18.99c	1.44 ± 0.04e	198.31d
	Y7	313.39 ± 9.50d	1.29 ± 0.01cd	242.94e
	Y9	212.09 ± 5.48b	1.36 ± 0.04de	155.57c
	Y10	120.29 ± 29.01a	1.03 ± 0.03a	116.79b
	Y12	319.72 ± 19.77d	1.20 ± 0.02bc	265.70f
	Y14	218.42 ± 18.99bc	1.16 ± 0.04b	187.76d
	2D55	303.89 ± 37.99d	1.27 ± 0.06cd	238.66e
Cellulase	N3	88.45 ± 7.45d	0.98 ± 0.07d	90.56d
	N7	27.14 ± 3.16ab	0.75 ± 0.04ab	36.03a
	N8	20.25 ± 5.47a	0.72 ± 0.02a	28.26a
	Y2	26.45 ± 5.47ab	0.74 ± 0.04a	35.91a
	Y5	40.92 ± 7.45b	0.85 ± 0.03bc	48.14b
	Y8	65.03 ± 5.20c	0.94 ± 0.03cd	68.93c
	2D55	169.04 ± 6.02e	1.43 ± 0.03e	118.21e

Note. Data represent mean ± SD. According to Tukey's test, values in the same column that are followed by different lowercase letters indicate significant differences at $P < 0.05$

conditions (60°C), exhibiting growth even in low pH environments down to pH 3. They also demonstrated moderate survivability at a high pH condition, up to pH 12. Strain Y7 exhibited 100% homology with *B. xiamenensis* strain MCCC 1A00008 (NR 148244.1), while strain 2D55 showed a 99% similarity to *B. licheniformis* strain 1-13AIA (FN 397486). Data for strain 2D55 has been previously published by Kazeem et al. (2016).

The evolutionary relationships between the enzyme-producing isolates and their closest type strains were analyzed using partial 16S rRNA gene sequences. A maximum-likelihood phylogenetic tree was constructed with the Tamura-Nei model after sequence alignment by ClustalW. As shown in Figure 1, the isolate *Bacillus xiamenensis* strain Y7 clustered closely with *B. xiamenensis* strain MCCC 1A00008, confirming its taxonomic placement among gram-positive rods.

Effect of pH and Temperature on the Hydrolase Activity

In the pour-plate assay, there were no indications of antagonistic interaction between strains 2D55 and Y7. Additionally, the liquid-based assay (Figure 2) revealed that the initial pH

Table 4
Cell morphology, colony morphology, and molecular identification of strain 2D55 and Y7

Characteristics	2D55	Y7
Cell morphology	Rod, Gram-positive	Rod, Gram-positive
Colony morphology	Whitish, rough, hair-like outgrowth, 2-3 mm in diameter	Large-sized, flat, irregular, Off-white, smooth colonies
Aerobic growth	+	+
Anaerobic growth	+ (facultative anaerobic)	+ (facultative anaerobic)
Growth at 30°C	++	++
Growth at 40°C	++	++
Growth at 50°C	+++	++
Growth at 60°C	++	++
Growth in pH 3	+	+
Growth in pH 5	++	++
Growth in pH 7	+++	+++
Growth in pH 9	+++	++
Growth in pH 12	+	+
Motility	+	+
Closest relatives in NCBI Genbank	*Bacillus licheniformis* strain 1-13AIA (FN 397486) *	*Bacillus xiamenensis* strain MCCC 1A00008 (NR 148244.1)
Percentage similarity	99%*	100%
Designated species name	*Bacillus licheniformis* 2D55*	*Bacillus xiamenensis* strain Y7
NCBI Accession no.	KT799651*	OK090764

Note. + = Positive reaction; – = Negative reaction; ++ = Moderate growth; +++ = High growth; NCBI = National Center for Biotechnology Information; * = Adapted from Kazeem et al. (2016)

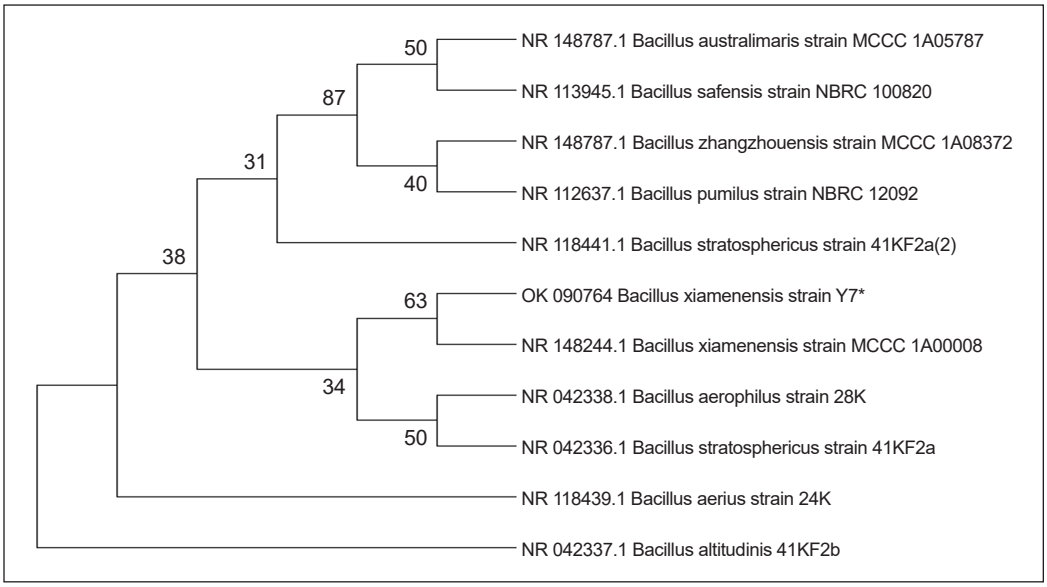


Figure 1. The strain names are preceded by the accession numbers of the type strains, enclosed in parentheses. The picture showed gram-positive rods of the *Bacillus xiamenensis* strain Y7

significantly influenced hydrolase production. In co-culture media, the highest hydrolase activities were observed at pH 7 ($P < 0.05$). While there was a notable decline in hydrolase production in co-culture at pH values below or above optimality, these productions surpassed those of both monocultures. Among these, amylase activity recorded the highest at 410.47 U/ml. Regarding monoculture, strain 2D55 consistently showed a notably higher activity ($P < 0.05$) than Y7 for amylase, lipase, and cellulase activity across all pH levels (5-9), except for protease activity. Upon closer examination, it was evident that the DS in total hydrolase activity was greater than the sum of all mono-culture activities, confirming the co-culture's hydrolase synergy.

Similarly, temperature had a significant impact on hydrolase activities, as shown in Figure 3. For amylase activity, both monocultures exhibited an increase from 35°C to optimal activity at 45°C, indicating the preferred temperature range for active amylase secretion. The co-culture maintained stable activity between 35 and 45°C, reaching its highest DS of 1.03 at 35°C. Strain 2D55 showed maximum protease activity at 35°C, decreasing at higher temperatures, while strain Y7 exhibited a gradual increase up to 45°C. However, the co-culture demonstrated significantly higher protease activities at all temperatures, peaking at 40°C with a DS value of 1.1. Both monocultures and the co-culture displayed optimal lipase activities at 40°C, with the co-culture exhibiting the highest activity at all temperature points, followed by strain 2D55 and Y7. Interestingly, contrary to the agar-based assay, strain Y7 exhibited cellulase activity in this liquid-based assay, albeit at a relatively low level. Meanwhile, strain 2D55 consistently showed higher

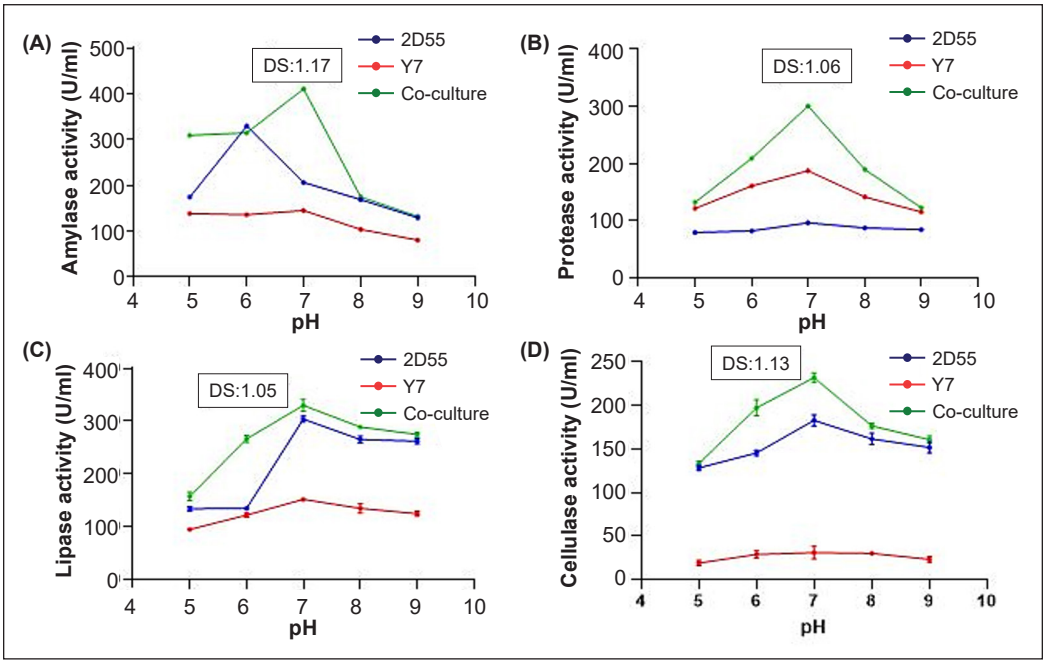


Figure 2. Effect of pH on the hydrolase activity of strain 2D55, Y7, and co-culture: (A) amylase, (B) protease, (C) lipase, and (D) cellulase
Note. DS = Degree of synergy; The standard errors (n = 3) are shown by error bars

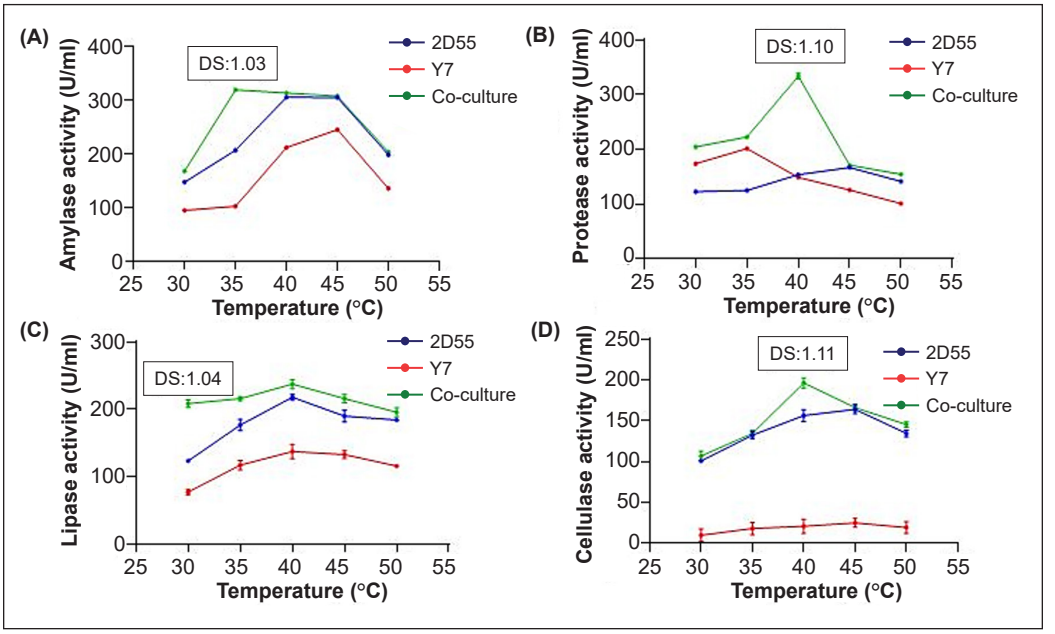


Figure 3. Effect of temperature on the hydrolase activity of strain 2D55, Y7, and co-culture: (A) amylase, (B) protease, (C) lipase, and (D) cellulase
Note. DS = Degree of synergy; The standard errors (n = 3) are shown by error bars

cellulase activity at all temperatures compared to strain Y7. The co-culture demonstrated similar cellulase activities to strain 2D55, reaching maximum synergy at 40°C with a DS value of 1.11.

In-vitro KW Biodegradation via SSF

To assess the effectiveness of the selected HPB in degrading KW, a 14-day SSF experiment was conducted, and the biological and physicochemical changes throughout the degradation process were monitored, as illustrated in Figure 4. In the monoculture treatment, the count of culturable bacteria steadily increased, reaching 11.40 log CFU/g by day 14 in sterile SSF. In contrast, the co-culture exhibited a dramatic increase, reaching 10.78 log CFU/g by day 6 and continuing to rise to 13.54 log CFU/g by day 14. In non-sterile SSF, bacterial growth was higher compared to sterile SSF, starting at 11.22 log CFU/g and reaching 14.00 log CFU/g by the end of the experiment for monocultures. However, the culturable bacteria in the non-sterile control showed the lowest count, reaching 13.70 log CFU/g by day 14. Surprisingly, the co-culture treatment in non-sterile SSF had the highest number of culturable bacteria, with bacterial growth reaching 11.95 log CFU/g by day 6 and continuing to increase to 14.60 log CFU/g by the end of the experiment (day 14). The pH values in both treatments ranged from 4.5 to 7.5. In sterile SSF, the pH shifted from acidic (pH 4) to neutral (pH 6), while in non-sterile SSF, the pH transitioned from neutral to slightly acidic, showing opposite trends compared to sterile SSF.

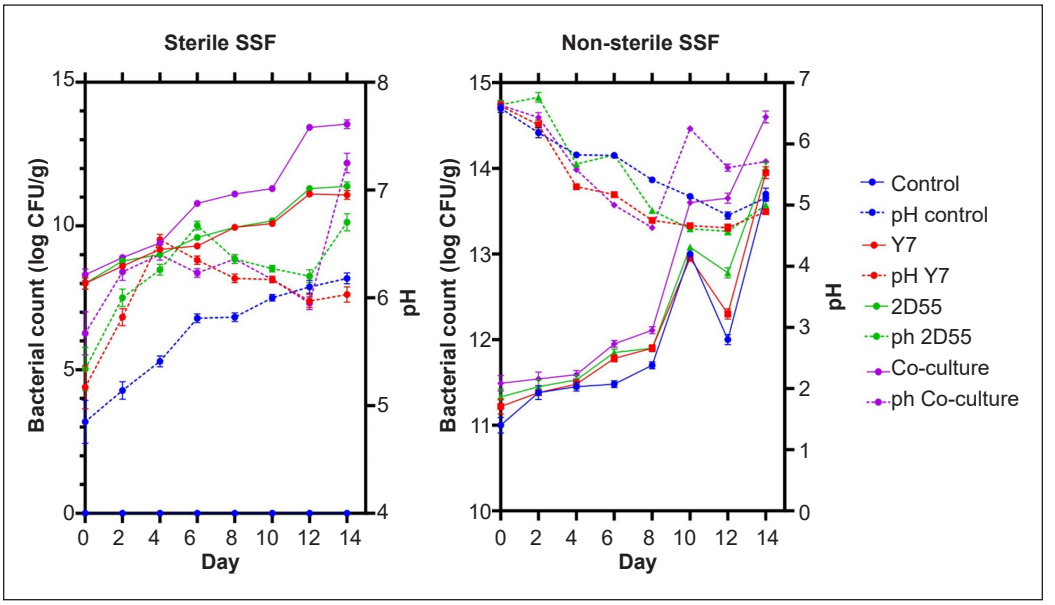


Figure 4. Bacterial count and the dynamics of pH during kitchen waste biodegradation via solid-state fermentation
Note. The standard errors (n = 3) are shown by error bars

The hydrolase activities observed in both sterile and non-sterile SSF are illustrated in Figure 5. In the inoculated flasks of both sterile and non-sterile SSF, amylase activity showed an overall increase up to day 10 of fermentation. Subsequently, amylase activity in sterile SSF began to decline towards day 14, while in non-sterile SSF, it exhibited a gradual increase during the same period. Protease activity in sterile SSF increased up to day 10, followed by a sudden decline towards the end of the experiment (day 14), whereas in non-sterile SSF, the decline started on day 8. In terms of lipase activity, the inoculated flasks in sterile SSF demonstrated an increase up to day 6, sharply declining after day 10. Conversely, in non-sterile SSF, lipase activity increased up to day 4 (except for strain Y7, where the increase was observed up to day 8) and gradually declined towards day 14. Cellulase activity in both sterile and non-sterile SSF increased up to day 12, followed by a gradual decline until day 14. The co-culture exhibited the highest amylase activity in sterile and non-sterile SSF, with optimal DS of 1.16 and 1.48, respectively. Similarly, for cellulase activity, the co-culture achieved the highest hydrolase activity with an optimal DS of 1.22 in sterile SSF and a DS of 1.03 in non-sterile SSF. Although the recorded lipase activity was comparatively lower than that of amylase and cellulase activity, the highest DS was found in lipase activity, with the co-culture achieving maximum synergy in both sterile and non-sterile SSF, with DS of 3.76 and 2.57, respectively.

A comprehensive analysis of the HPB's capacity to degrade KW was conducted based on the remaining TS and GDR, as illustrated in Figure 6. It is evident that inoculated KW exhibited higher degradation rates compared to the uninoculated control in both sterile and non-sterile SSF (Figures 6A and 6B). In sterile SSF, the TS degradation of the co-culture significantly increased up to 39% on day 2, reaching 54% by the end of the experiment (day 14). Similarly, monoculture treatments achieved slightly lower percentages, with Y7 and 2D55 reaching 48 and 51%, respectively, on day 14. In non-sterile SSF, the TS degradation of the co-culture also showed a significant increase, reaching 44% on day 2 and 65% by the end of the experiment, outperforming monoculture treatments, where Y7 and 2D55 achieved 49 and 53%, respectively. Co-culture inoculation resulted in significantly higher TS degradation compared to monoculture and uninoculated control ($P < 0.05$). Regarding GDR, co-culture inoculation in sterile SSF achieved a high percentage of 7% on day 2, rapidly increasing to 25% by the end of the experiment. In contrast, monoculture treatments reached lower percentages, with Y7 and 2D55 achieving 15 and 17%, respectively, on the final day. In non-sterile SSF, co-culture inoculation also exhibited a rapid increase in GDR, reaching 41%, whereas Y7 and 2D55 monoculture treatments achieved only 20 and 23%, respectively, by the end of the experiment. Co-culture inoculation resulted in significantly higher GDR percentages compared to monoculture and uninoculated control ($P < 0.05$). These findings demonstrate that co-culture inoculation in both sterile and non-sterile SSF can effectively enhance TS and GDR degradation of KW.

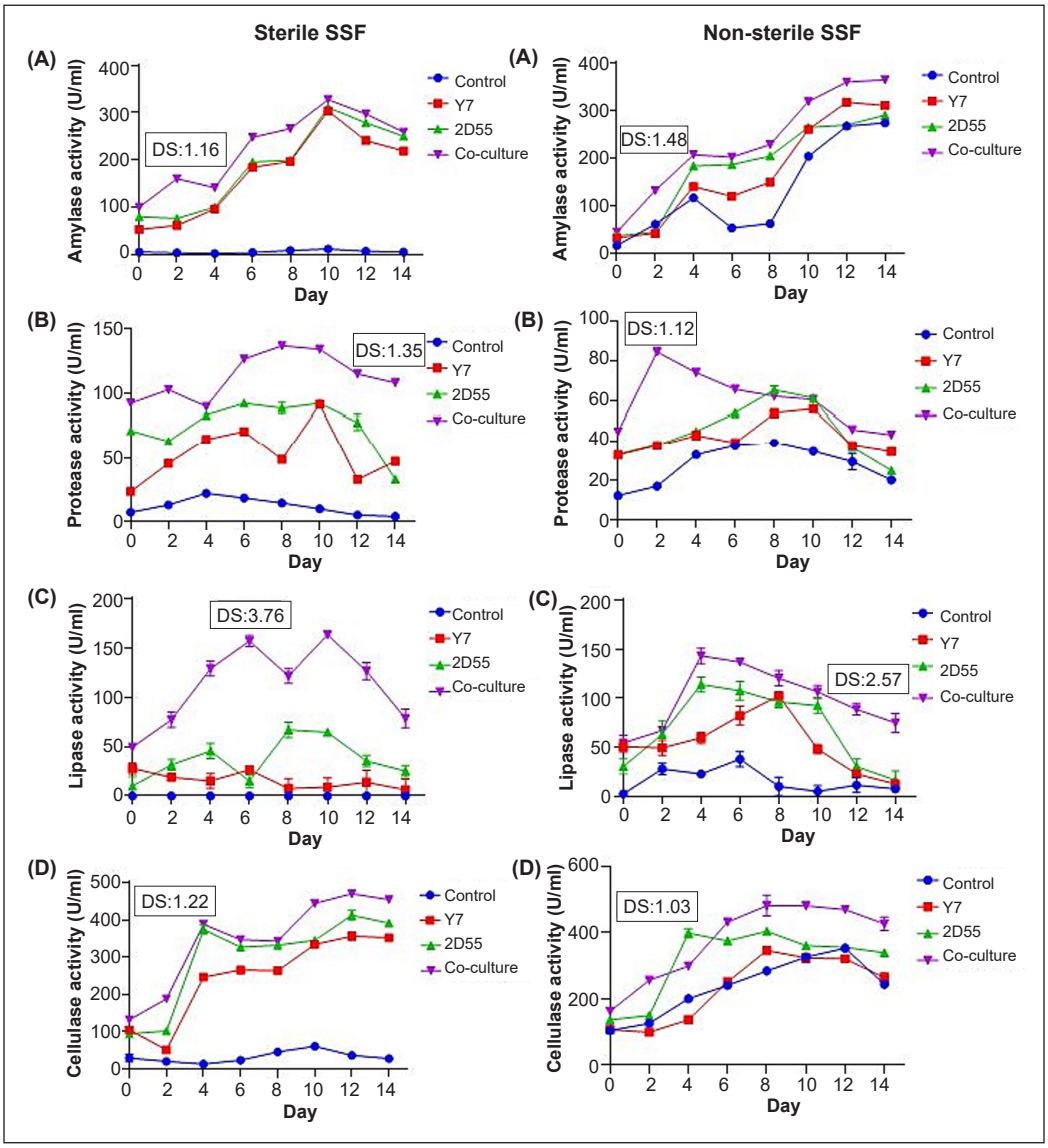


Figure 5. Hydrolase activity of control, strain 2D55, Y7, and co-culture in sterile and non-sterile solid-state fermentation: (A) amylase, (B) protease, (C) lipase, and (D) cellulase
Note. DS = Degree of synergy; The standard errors (n = 3) are shown by error bars

Pearson's correlation coefficients were computed to assess the relationships between hydrolase activities, bacterial count, TS, and GDR (Table 5). Strong correlations (**$P < 0.01$) were observed between bacterial count and TS, GDR, and all hydrolase activities in sterile SSF. Conversely, negative correlations were noted between bacterial count and protease and lipase activity. In sterile SSF, robust correlations (**$P < 0.01$ and *$P < 0.05$) were found between TS and all hydrolase activities. However, in non-sterile

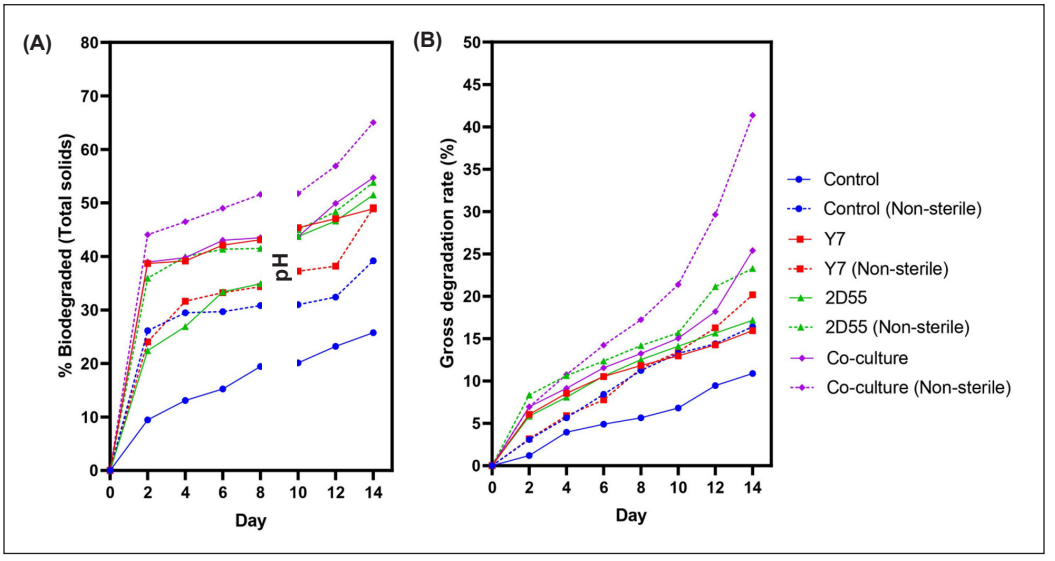


Figure 6. Effect of different hydrolase-producing bacteria on kitchen waste biodegradation. (A) Percentage of biodegradation based on total solids left within 14-day solid-state fermentation (SSF); (B) Gross degradation rate of KW within 14-day SSF

Table 5
Pearson's correlation coefficient between hydrolase activity, bacterial count, total solids (TS), and gross degradation rate (GDR)

	Bacterial count	TS	GDR	Amylase	Protease	Lipase	Cellulase
Sterile							
Bacterial count	1	0.691**	0.619**	0.861**	0.777**	0.572**	0.865**
TS		1	0.907**	0.797**	0.509**	0.432*	0.798**
GDR			1	0.777**	0.453**	0.403*	0.768**
Non-sterile							
Bacterial count	1	0.675**	0.875**	0.853**	-0.071	-0.071	0.568**
TS		1	0.850**	0.806**	0.434*	0.168	0.783**
GDR			1	0.885**	0.161	0.402*	0.768**

Note. Significant correlations ($P < 0.01$) are marked with ** or ($P < 0.05$) with *

SSF, no significant correlations were observed between TS and lipase activity. Similar trends were observed for GDR, with strong correlations identified between GDR and all hydrolase activities in sterile SSF. In contrast, no significant correlation was found between GDR and protease activity in non-sterile SSF.

Analysis of the Composting Process

The kinetics of KW degradation during composting were elucidated by the observed hydrolase activity (Figure 7). In the inoculated compost, amylase activity exhibited an

overall increase until week 4 of composting, followed by a decline starting from week 5. Conversely, the control compost showed a slight increase in amylase activity during the first two weeks of composting, followed by a decline until the end of the process. Cellulase activity in the inoculated compost decreased after week four, mirroring the trend observed in amylase activity, while cellulase activity in the control compost began to decline after week two. For protease activity, the inoculated compost displayed a decline in the first two weeks, followed by a slight increase in week 3, whereas the control compost showed relatively stable protease activity until a slight increase towards the end of the composting period. Inoculated compost exhibited an increase in lipase activity until week 6, followed by a gradual decline towards the end of the experiment (week 8). Conversely, lipase activity in the control compost started declining after week 5. All hydrolase activities in the inoculated compost were significantly higher compared to the control compost ($P < 0.05$).

Figure 8 illustrates the changes in physicochemical and biological parameters during the composting process. In both inoculated and control composts, the moisture content remained within the range of 40 to 70% throughout the experiment (Figure 8A). The temperature profiles (Figure 8B) exhibited a similar trend in composting processes with and without HPB. There was a notable rise ($P < 0.05$) in temperature during the initial days of composting, followed by a gradual decrease after reaching the peak (5-10 days). Subsequently, the temperature stabilized between 24 and 30°C after 30 days, approximating

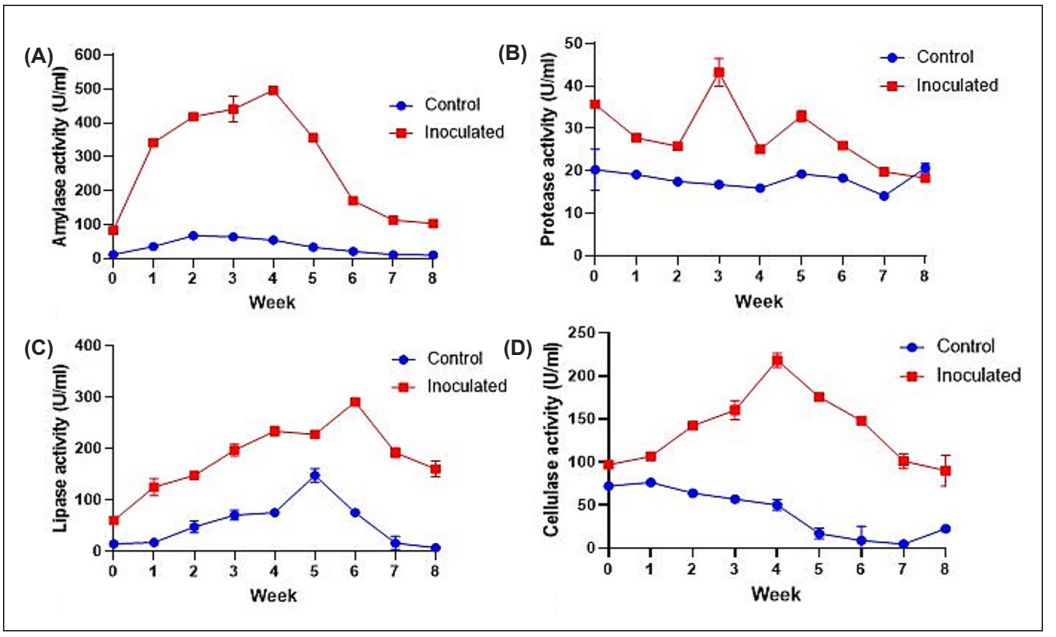


Figure 7. Hydrolase activity of inoculated and control compost during the composting process: (A) amylase, (B) protease, (C) lipase, and (D) cellulase
Note. The standard errors (n = 3) are shown by error bars

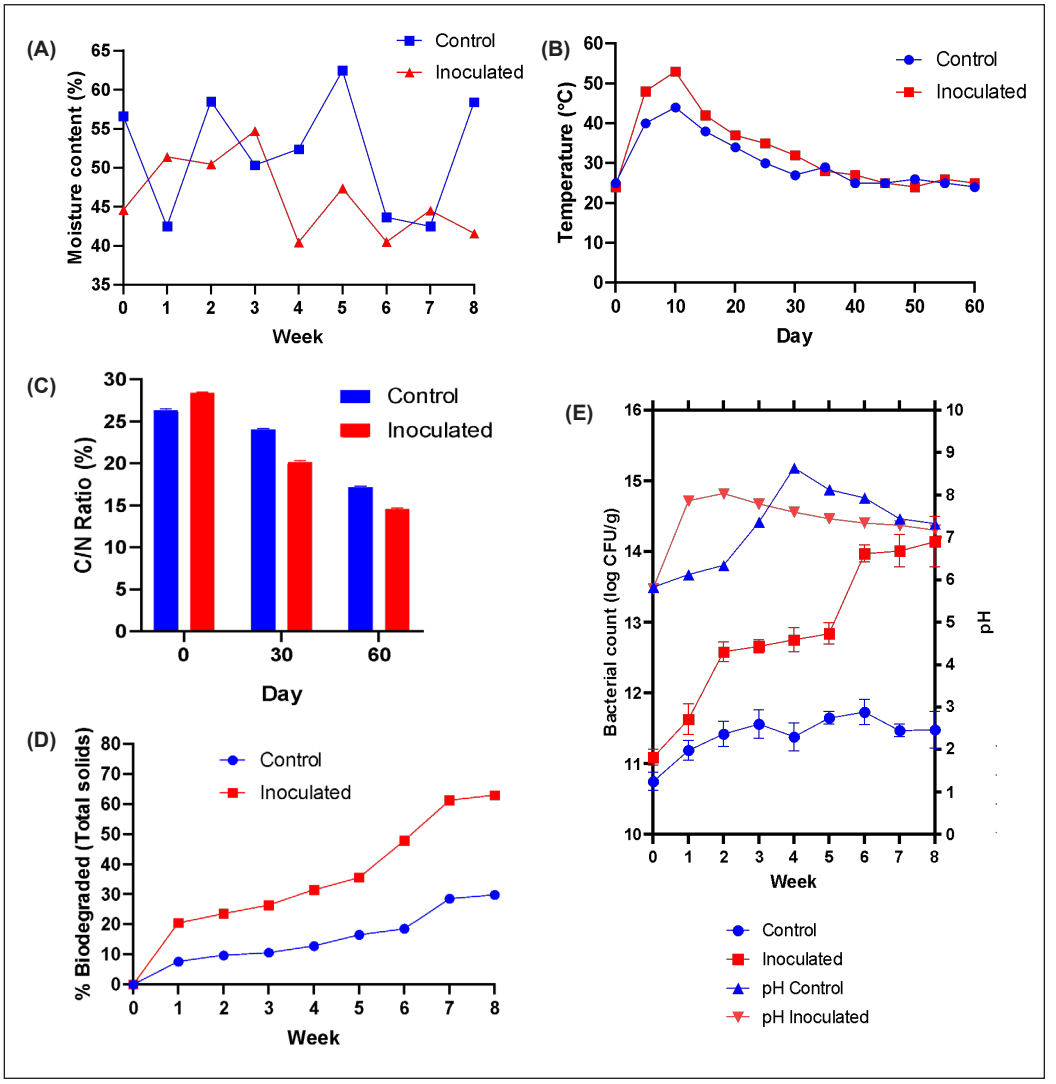


Figure 8. Biological and physico-chemical changes during the composting process: (A) moisture content, (B) temperature profile, (C) C/N ratio, (D) percentage of biodegradation based on the amount of total solids that remain, and (E) bacterial population and pH dynamics
Note. The standard errors (n = 3) are shown by error bars

room temperature. The carbon-to-nitrogen (C/N) ratio significantly decreased towards the end of the composting process in both inoculated and control composts (Figure 8C). The C/N ratio of inoculated compost (14.58) was substantially lower ($P < 0.05$) than the control compost (17.23). Both trials exhibited a final reduction in the C/N ratio ranging from 30 to 48%.

Additionally, the TS degradation in inoculated compost increased drastically up to 21% in the first week and continued to rise to 63% by the end of the experiment. In contrast, TS

degradation in the control compost was slightly lower, with the highest percentage achieved being 29% by the eighth week. In terms of bacterial growth, culturable bacteria steadily increased in the control compost, reaching 11.56 log CFU/g by the third week (Figure 8E). In contrast, the inoculated compost exhibited a drastic increase, reaching 12.58 log CFU/g by the second week and continuing to rise to 13.97 log CFU/g by the sixth week. The highest bacterial count was observed in the inoculated compost, reaching 14.14 log CFU/g by the end of analysis. The bacterial growth in the control compost was lower, recording 11.48 log CFU/g by the end of analysis. The pH values in both composts varied from 5.0 to 8.5, transitioning from acidic (pH 5) to neutral (pH 7) and weakly alkaline, demonstrating the organic matter's stability.

Table 6 presents Pearson's correlation coefficients calculated for various parameters, including hydrolase activities, C/N ratio, bacterial count, TS, and pH. Strong positive correlations (**$P < 0.01$) were observed between bacterial count and TS in both the inoculated and control composts. Conversely, negative correlations were found between the C/N ratio and bacterial count, as well as between the C/N ratio and TS in both compost types. Additionally, positive correlations were noted between pH and bacterial count, as well as between pH and TS. In the inoculated compost, strong correlations were discovered (**$P < 0.01$ and *$P < 0.05$) between lipase activity and bacterial count, hydrolase activity and TS, and TS and C/N ratio. However, no significant correlations were observed between protease activity and bacterial count, pH, or C/N ratio in the control compost. Similarly, there were no noteworthy correlations discovered between amylase activity and other parameters in both compost types, except for pH and TS. In the inoculated compost, a significant correlation (**$P < 0.05$) was found between cellulase activity and pH. In

Table 6
Pearson's correlation coefficient between hydrolase activity, carbon/nitrogen (C/N), bacterial count, total solids (TS), and pH

	C/N	Bacterial count	TS	pH	Amylase	Protease	Lipase	Cellulase
Inoculated								
C/N	1	-0.983**	-0.994**	-0.795*	-0.148	0.997**	-0.655	-0.057
Bacterial count		1	0.946**	0.248	-0.262	-0.586**	0.672**	0.000
TS			1	0.261	0.234*	0.689**	0.564**	0.113
pH				1	0.703**	-0.140	0.447*	0.413*
Control								
C/N	1	-0.663	-0.980**	-0.275	0.307	-0.249	0.368	0.965**
Bacterial count		1	0.539**	0.603**	0.171	-0.124	0.502**	-0.560*
TS			1	0.517**	0.416	0.210	0.058	0.831*
pH				1	0.087	-0.316	0.625*	-0.637*

Note. Significant correlations ($P < 0.01$) are marked with ** or ($P < 0.05$) with *

contrast, in the control compost, a significant correlation (**$P < 0.01$) was observed between cellulase activity and C/N ratio.

Table 7 summarizes the characteristics of composts made without HPB (control), including the suggested ranges for determining compost maturity and stability. After 8 weeks, both composts were well-matured, exhibiting a germination index (>100%), a pH of around 7, a C/N ratio of less than 20, a dark brown color, and an earthy smell. Both composts also showed non-phytotoxicity, as indicated by their pH, C/N ratio, and high germination index.

Table 7
Properties of final composts in week 8 and the typical range

Properties	Inoculated compost	Control	Typical range
GI (%)	178.25 ± 23.2	176.01 ± 31.6	> 100% (Palaniveloo et al., 2020)
C/N ratio	14.58 ± 0.14	17.23 ± 0.09	< 20 (Onwosi et al., 2017)
pH	7.17 ± 0.08	7.32 ± 0.04	6-8 (Varma et al., 2017)
Colour	Dark brown	Dark brown	Dark brown (Khalil et al., 2021)
Smell	Earthy smell (Start from week 3)	Earthy smell (Start week 7)	Earthy smell (Hubbe et al., 2010)

Note. Data represent mean ± SD; GI = Germination index; C/N ratio = Carbon/nitrogen ratio

DISCUSSION

In this study, SSF was utilized to assess the potential of various HPB to degrade starch, protein, fat, and total sugars, thereby enhancing KW biodegradation efficiency. After primary isolation and screening of HPB, two isolates, namely strain 2D55 and Y7, were selected. Molecular analysis confirmed that strain 2D55 belonged to the *B. licheniformis* species, whereas strain Y7 belonged to the *B. xiamenensis* phylogeny. *Bacillus licheniformis* is commonly utilized in the pharmaceutical industry for the production of biochemicals, enzymes, antibiotics, and aminopeptidase (Lee et al., 2017). On the other hand, *B. xiamenensis* was initially discovered in the intestines of a flathead mullet on Xiamen Island, China (Lai et al., 2014). According to Amna et al. (2020), this species exhibits plant growth-promoting capabilities, suggesting its potential to enhance plant development and act as a biological control against phytopathogens. While information on these species is limited, *B. xiamenensis* has not been previously associated with KW biodegradation capacity, and no research on the synergistic effect of both strains on KW degradation rate currently exists. This study aims to address this knowledge gap and explore the combined potential of these HPB strains in enhancing KW degradation. The biocompatibility assay confirmed that the selected strains, *B. licheniformis* 2D55 and *B. xiamenensis* Y7, demonstrated no inhibitory interactions when tested against each other. No inhibition zones exceeding 4 mm in diameter were observed for any of

the strains during the 3-day incubation period, indicating mutual compatibility. These results validated the feasibility of co-culturing these strains for subsequent experiments in the study.

In the quantitative assay, it was intriguing to observe that the hydrolase activity of the HPB co-culture surpassed that of the corresponding monocultures, as evidenced by the significantly higher DS value ($P < 0.05$). This observation aligns with previous studies where co-cultures of *B. velezensis* and *B. paralicheniformis* exhibited greater reductions in various components of FW, including total solids, fat, sugar, reducing sugar, protein, and starch, compared to monocultures (Roslan et al., 2021). Similarly, another study reported that a bacterial co-culture was able to eliminate up to 98% of 17-estradiol, a common contaminant in water and soil, within 7 days of incubation, whereas monoculture achieved only 77% removal (M. Li et al., 2018). These findings suggest that microbial symbiotic relationships during fermentation may play a pivotal role in generating such significant synergistic effects.

Throughout the 14-day SSF process for KW biodegradation, bacterial proliferation was highly active, especially in the co-culture treatment, where the total culturable bacteria reached the highest level of 14.6 log CFU/g in the non-sterile SSF system (Figure 4). The decrease in bacterial population on day 12 for monocultures and control under non-sterile conditions could be due to the available nutrients in the substrate being significantly depleted, especially under non-sterile conditions, where native microbial communities compete with the inoculated bacteria for resources. This competition could lead to a temporary decline in the population of the inoculated strains as they experience nutrient limitations. Monocultures and control do not benefit from the synergistic interactions present in the co-culture, making them more susceptible to nutrient depletion or the inhibitory effects of accumulated metabolic byproducts (Bernal et al., 2009). The pH levels in both SSF systems fluctuated between 4.5 and 6.5 over the 14-day fermentation period. According to Chen et al. (2015), optimal pH values for degradation typically range from 5.5 to 8, facilitating microbial decomposition of organic materials. The initial decrease in pH during the starting point of degradation may be attributed to the volatilization of early ammonia and the release of organic acids during the degradation of simple organic substrates (Azim et al., 2018). However, as the co-culture consumes these acids or as their production declines in the later stages of fermentation, the buffering effect diminishes, allowing the pH to increase. Under sterile conditions, the absence of competing microbial populations that might utilize ammonia or other alkaline compounds could further amplify this effect (Zhang & Sun, 2017).

The unexpected trends in the pH and temperature optima for the enzymes in the co-culture system may be attributed to several factors related to microbial interactions and environmental dynamics within the co-culture. In SSF, localized microenvironments within

the solid substrate can lead to variations in pH and temperature gradients, particularly in a co-culture system where the metabolic activities of multiple strains interact. These microenvironmental factors may explain why the co-culture exhibited different enzymatic optima compared to the monocultures. In the co-culture, the interaction between *B. licheniformis* 2D55 and *B. xiamenensis* Y7 may have led to changes in the extracellular environment, such as shifts in pH or the production of secondary metabolites. These interactions could modulate enzyme activity and stability, potentially altering the observed optima for enzymatic functions. Previous studies have shown that co-cultures often produce synergistic effects that enhance enzymatic activity or change the optimal conditions for enzyme performance (Cerda et al., 2017). Therefore, careful selection of appropriate members for co-culture systems is crucial for establishing an effective long-term waste disposal strategy (Mujtaba & Lee, 2016).

Our findings indicate that during the initial phases of fermentation, amylase and cellulase exhibited significant activity in both sterile and non-sterile SSF co-culture inoculations, aligning with prior research documenting elevated levels of these enzymes at the onset of decomposition (Fan et al., 2018; Khalil et al., 2021). Conversely, protease and lipase displayed comparatively lower activity, possibly due to the predominance of rice and vegetable components in KW, which necessitate greater amylase and cellulase activity for breakdown. Compared to monoculture treatments, both sterile and non-sterile SSF demonstrated the highest percentage of biodegraded TS, reaching up to 54.73 and 65%, respectively, showing that co-culture inoculation has significant synergistic effects on TS degradation. These findings resonate with recent reports by Roslan et al. (2021), showing a reduction of FW solid composition by over 60% with *Bacillus* spp. co-culture inoculation, and another study reporting a 23.92% reduction in FW mass with microbial consortium addition (Zhao et al., 2017). The observed synergy in the co-culture fermentation of *Bacillus* spp. can be attributed to the complementary enzymatic activities provided by each strain. *Bacillus licheniformis* 2D55 contributes strong cellulase and lipase activities, while *B. xiamenensis* Y7 enhances protease and amylase activity, together forming a robust enzymatic profile capable of degrading a range of complex organic compounds found in kitchen waste. This complementarity allows for a more thorough breakdown of proteins, carbohydrates, lipids, and cellulose compared to monoculture treatments, where a single strain's enzymatic activity is more limited (Kapoore et al., 2021).

Co-culture treatments also exhibited the highest GDR percentages for both sterile (25.39%) and non-sterile (41.36%) SSF, underscoring the efficacy of co-culture inoculation in enhancing KW degradation and biomass reduction ($P < 0.05$). The observed difference in GDR between sterile and non-sterile conditions may be attributed to the role of native microbial communities in the non-sterile environment. Under non-sterile conditions, indigenous microorganisms interact with the introduced *Bacillus* spp. strains, potentially

creating synergistic effects that enhance enzymatic activity and accelerate organic matter breakdown (Niu & Li, 2022). This additional microbial contribution likely accounts for the higher GDR observed in non-sterile SSF compared to sterile conditions, where only the *Bacillus* spp. co-culture was active. The positive correlation between GDR and hydrolase activities suggests that robust hydrolase activity accelerates KW degradation, potentially shortening the initial lag time of the biological process (Rastogi et al., 2020). Overall, these results highlight the potential of the isolated HPB in the KW degradation process, showcasing their ability to effectively break down KW and reduce its biomass.

During the composting of KW, the hydrolase activities exhibited a similar trend to those observed in SSF, with amylase and cellulase displaying the highest activity during the initial stages of composting, followed by protease and lipase. Both the inoculated and control composts maintained moisture levels ranging from 40 to 70%, ideal for fostering biological activity during composting (Abdel-Rahman et al., 2016). However, unlike the inoculated compost, the control compost lacks the addition of HPB, which enhances microbial activity and regulates the degradation process. Metabolic heat generated by inoculated microbes during active degradation can facilitate consistent evaporation and maintain a more stable moisture balance. In the absence of these inoculants, the control compost becomes more susceptible to moisture fluctuations caused by uneven degradation or environmental factors. Studies have shown that inoculated compost tends to exhibit more uniform moisture distribution due to enhanced microbial metabolism and substrate breakdown (Zhang & Sun, 2017). Throughout the thermophilic phase of composting, the inoculated compost notably exhibited a more rapid temperature rise to thermophilic levels (> 45°C) in the early stages of composting compared to the control. Moreover, the thermophilic phase in the inoculated compost persisted for a longer duration, spanning from day 5 to day 10, with temperatures ranging from 48 to 53°C, respectively. These observations suggest that the inoculated compost enhanced microbial population and activity during composting. The elevated temperatures are believed to result from increased biological activity, stemming from the breakdown of carbon bonds during the biodegradation process (Azim et al., 2018).

Both treatments exhibited a decrease in the C/N ratio as composting progressed, with the inoculated compost showing a higher reduction in C/N ratio (48.71%) compared to the control (34.73%). Research indicates that as waste gradually degrades, the C/N ratio tends to decrease (Onwosi et al., 2017). The inoculated compost demonstrated a significantly higher percentage of biodegraded TS ($P < 0.05$), reaching 63.02%, in contrast to the control, which achieved 29.80%. The positive correlation observed between TS and hydrolase activities suggests that robust hydrolase activities facilitate KW degradation. A similar trend was noted between the bacterial count in the co-culture SSF and that in the inoculated compost, with the highest culturable bacteria (14.14 log CFU/g) recorded

throughout the two-month composting process. In both composts, the pH fluctuated from acidic to neutral to mildly alkaline toward the end of the process. The initial acidity of the pH in the early stages of composting may be attributed to the breakdown of simple organic compounds and the initial ammonia's volatilization, resulting in the formation of organic acids. Subsequently, the pH became alkaline due to mineralization and the decomposition of quickly biodegradable organic components (Bashir et al., 2021).

The properties of the final composts were evaluated, with a particular focus on odor emission, a common issue in composting processes. Typically, the rapid transition from acidic to alkaline conditions initially results in unpleasant odors. However, in our study, the inoculated compost began emitting an earthy odor after only three weeks, whereas the control compost took six weeks to achieve the same. Both samples exhibited values exceeding 100% for non-phytotoxicity, with the inoculated compost's germination index slightly higher than that of the control. Research suggests that substances with GI values below 50% are highly phytotoxic, those between 50% and 80% are moderately phytotoxic, values above 80% are non-phytotoxic, and values over 100% have a phytonutrient effect (Siles-Castellano et al., 2020).

CONCLUSION

The co-culture of *B. licheniformis* 2D55 and *B. xiamenensis* Y7 exhibited heightened hydrolase activities and demonstrated remarkable synergism, surpassing the levels observed in monocultures. These findings imply that employing HPB in composting accelerates vegetable waste (VW) breakdown, enhances hydrolase activity, yields compost with an earthy scent, and mitigates phytotoxicity. While this study highlights the enhanced hydrolase activity observed in co-culture conditions, the exact biological mechanisms driving this synergy remain to be explored. Detailed molecular studies, such as transcriptomics or metabolomics, could provide insight into these specific pathways and metabolic exchanges, offering a deeper understanding of the cooperative dynamics that enhance biodegradation rates. Large-scale trials are also necessary to assess the practical feasibility and economic viability of employing these bacterial co-cultures in waste management. Besides, future research should assess the efficacy of these *Bacillus* spp. co-cultures on a broader spectrum of organic waste types, including higher-protein and lipid-rich waste, to ensure wider applicability. Such insights could pave the way for developing a valuable tool to offer an alternative environmental approach and enhance waste treatment options holistically.

ACKNOWLEDGEMENTS

This work was supported by the Putra Graduate Initiative Grant (9653800) allocated by UPM, Malaysia.

REFERENCES

Abdel-Rahman, M. A., El-Din, M. N., Refaat, B. M., Abdel-Shakour, E. H., Ewais, E. E.-D., & Alrefaey, H. M. A. (2016). Biotechnological application of thermotolerant cellulose-decomposing bacteria in composting of rice straw. *Annals of Agricultural Sciences*, *61*(1), 135–143. https://doi.org/10.1016/j.aoas.2015.11.006

Abubakar, I. R., Maniruzzaman, K. M., Dano, U. L., AlShihri, F. S., AlShammari, M. S., Ahmed, S. M. S., Al-Gehlani, W. A. G., & Alrawaf, T. I. (2022). Environmental sustainability impacts of solid waste management practices in the Global South. *International Journal of Environmental Research and Public Health*, *19*(19), 12717. https://doi.org/10.3390/IJERPH191912717

Amna., Xia, Y., Farooq, M. A., Javed, M. T., Kamran, M. A., Mukhtar, T., Ali, J., Tabassum, T., ur Rehman, S., Hussain Munis, M. F., Sultan, T., & Chaudhary, H. J. (2020). Multi-stress tolerant PGPR *Bacillus xiamenensis* PM14 activating sugarcane (*Saccharum officinarum* L.) red rot disease resistance. *Plant Physiology and Biochemistry*, *151*, 640–649. https://doi.org/10.1016/J.PLAPHY.2020.04.016

Azim, K., Soudi, B., Boukhari, S., Perissol, C., Roussos, S., & Thami Alami, I. (2018). Composting parameters and compost quality: A literature review. *Organic Agriculture*, *8*, 141–158. https://doi.org/10.1007/s13165-017-0180-z

Bashir, S., Gulshan, A. B., Iqbal, J., Husain, A., Alwahibi, M. S., Alkahtani, J., Dwiningsih, Y., Bakhsh, A., Ahmed, N., Khan, M. J., Ibrahim, M., & Diao, Z.-H. (2021). Comparative role of animal manure and vegetable waste induced compost for polluted soil restoration and maize growth. *Saudi Journal of Biological Sciences*, *28*(4), 2534–2539. https://doi.org/10.1016/J.SJBS.2021.01.057

Bernal, M. P., Alburquerque, J. A., & Moral, R. (2009). Composting of animal manures and chemical criteria for compost maturity assessment. A review. *Bioresource Technology*, *100*(22), 5444–5453. https://doi.org/10.1016/J.BIORTECH.2008.11.027

Cerda, A., Gea, T., Vargas-García, M. C., & Sánchez, A. (2017). Towards a competitive solid state fermentation: Cellulases production from coffee husk by sequential batch operation and role of microbial diversity. *Science of the Total Environment*, *589*, 56–65. https://doi.org/10.1016/j.scitotenv.2017.02.184

Chen, R., Wang, Y., Wang, W., Wei, S., Jing, Z., & Lin, X. (2015). N_2O emissions and nitrogen transformation during windrow composting of dairy manure. *Journal of Environmental Management*, *160*, 121–127. https://doi.org/10.1016/J.JENVMAN.2015.06.021

Drissi Kaitouni, L. B., Anissi, J., Sendide, K., & El Hassouni, M. (2020). Diversity of hydrolase-producing halophilic bacteria and evaluation of their enzymatic activities in submerged cultures. *Annals of Microbiology*, *70*, 33. https://doi.org/10.1186/s13213-020-01570-z

Fan, Y. V., Lee, C. T., Klemeš, J. J., Chua, L. S., Sarmidi, M. R., & Leow, C. W. (2018). Evaluation of Effective Microorganisms on home scale organic waste composting. *Journal of Environmental Management*, *216*, 41–48. https://doi.org/10.1016/j.jenvman.2017.04.019

Grossart, H.-P., Schlingloff, A., Bernhard, M., Simon, M., & Brinkhoff, T. (2004). Antagonistic activity of bacteria isolated from organic aggregates of the German Wadden Sea. *FEMS Microbiology Ecology*, *47*(3), 387–396. https://doi.org/10.1016/S0168-6496(03)00305-2

Hubbe, M. A., Nazhad, M., & Sánchez, C. (2010). Composting as a way to convert cellulosic biomass and organic waste into high-value soil amendments: A review. *BioResources*, *5*(4), 2808–2854. https://doi.org/10.15376/BIORES.5.4.2808-2854

Kaewlaoyoong, A., Cheng, C.-Y., Lin, C., Chen, J.-R., Huang, W.-Y., & Sriprom, P. (2020). White rot fungus *Pleurotus pulmonarius* enhanced bioremediation of highly PCDD/F-contaminated field soil via solid state fermentation. *Science of The Total Environment*, *738*, 139670. https://doi.org/10.1016/J.SCITOTENV.2020.139670

Kapoore, R. V., Padmaperuma, G., Maneein, S., & Vaidyanathan, S. (2021). Co-culturing microbial consortia: Approaches for applications in biomanufacturing and bioprocessing. *Critical Reviews in Biotechnology*, *42*(1), 46–72. https://doi.org/10.1080/07388551.2021.1921691

Kazeem, M. O., Shah, U. K. M., Baharuddin, A. S., & Rahman, N. A. (2016). Enhanced cellulase production by a novel thermophilic *Bacillus licheniformis* 2D55: Characterization and application in lignocellulosic saccharification. *BioResources*, *11*(2), 5404–5423. https://doi.org/10.15376/BIORES.11.2.5404-5423

Khalil, A. I., Ahmed, M. N., & Badry, H. H. (2021). Monitoring the changes of some physical, chemical, microbial and enzymatic parameters during composting of three garden wastes. *Alexandria Science Exchange Journal*, *42*(1), 223–237. https://doi.org/10.21608/ASEJAIQJSAE.2021.160006

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/MOLBEV/MSY096

Kutsanedzie, F., Rockson, G. N. K., Aklaku, E. D., & Achio, S. (2012). Comparisons of compost maturity indicators for two field scale composting systems. *International Research Journal of Applied and Basic Sciences*, *3*(4), 713–720.

Lai, Q., Liu, Y., & Shao, Z. (2014). *Bacillus xiamenensis* sp. nov., isolated from intestinal tract contents of a flathead mullet (*Mugil cephalus*). *Antonie van Leeuwenhoek*, *105*, 99–107. https://doi.org/10.1007/S10482-013-0057-4

Lee, C., Kim, J. Y., Song, H. S., Kim, Y. B., Choi, Y.-E., Yoon, C., Nam, Y.-D., & Roh, S. W. (2017). Genomic analysis of *Bacillus licheniformis* CBA7126 isolated from a human fecal sample. *Frontiers in Pharmacology*, *8*, 724. https://doi.org/10.3389/FPHAR.2017.00724

Leege, P. B. (1998). Introduction of test methods for the examination of composting and compost. In S. Brown, J. S., Angle, & L. Jacobs (Eds.), *Beneficial co-utilization of agricultural, municipal and industrial by-products* (pp. 269–282). Springer. https://doi.org/10.1007/978-94-011-5068-2_22

Leow, C. W., Fan, Y. V., Chua, L. S., Muhamad, I. I., Klemeš, J. J., & Lee, C. T. (2018). A review on application of microorganisms for organic waste management. *Chemical Engineering Transactions*, *63*, 85–90. https://doi.org/10.3303/CET1863015

Li, M., Zhao, X., Zhang, X., Wu, D., & Leng, S. (2018). Biodegradation of 17β-estradiol by bacterial co-culture isolated from manure. *Scientific Reports*, *8*, 3787. https://doi.org/10.1038/S41598-018-22169-0

Li, T., Liu, L., Wei, X., Zhang, H., & Fang, P. (2014). Comparison of the food waste degradation rate and microbial community in the matrix between two biodegradation agents in a food waste disposer. *Waste Management*, *34*(12), 2641–2646. https://doi.org/10.1016/J.WASMAN.2014.08.020

Miller, G. L. (1959). Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Analytical Chemistry*, *31*(3), 426–428. https://doi.org/10.1021/AC60147A030

Mujtaba, G., & Lee, K. (2016). Advanced treatment of wastewater using symbiotic co-culture of microalgae and bacteria. *Applied Chemistry for Engineering*, *27*(1), 1–9. https://doi.org/10.14478/ACE.2016.1002

Niu, J., & Li, X. (2022). Effects of microbial inoculation with different indigenous *Bacillus* species on physicochemical characteristics and bacterial succession during short-term composting. *Fermentation*, *8*(4), 152. https://doi.org/10.3390/FERMENTATION8040152

Onwosi, C. O., Igbokwe, V. C., Odimba, J. N., Eke, I. E., Nwankwoala, M. O., Iroh, I. N., & Ezeogu, L. I. (2017). Composting technology in waste stabilization: On the methods, challenges and future prospects. *Journal of Environmental Management*, *190*, 140–157. https://doi.org/10.1016/j.jenvman.2016.12.051

Palaniveloo, K., Amran, M. A., Norhashim, N. A., Mohamad-Fauzi, N., Peng-Hui, F., Hui-Wen, L., Kai-Lin, Y., Jiale, L., Chian-Yee, M. G., Jing-Yi, L., Gunasekaran, B., & Razak, S. A. (2020). Food waste composting and microbial community structure profiling. *Processes*, *8*(6), 723. https://doi.org/10.3390/pr8060723

Privé, F., Newbold, C. J., Kaderbhai, N. N., Girdwood, S. G., Golyshina, O. V., Golyshin, P. N., Scollan, N. D., & Huws, S. A. (2015). Isolation and characterization of novel lipases/esterases from a bovine rumen metagenome. *Applied Microbiology and Biotechnology*, *99*, 5475–5485. https://doi.org/10.1007/s00253-014-6355-6

Rastogi, M., Nandal, M., & Khosla, B. (2020). Microbes as vital additives for solid waste composting. *Heliyon*, *6*(2), e03343. https://doi.org/10.1016/j.heliyon.2020.e03343

Roslan, M. A. M., Jefri, N. Q. U. A., Ramlee, N., Rahman, N. A. A., Chong, N. H. H., Bunawan, H., Bharudin, I., Kadir, M. H. A., Mohammad, M., & Razali, H. (2021). Enhancing food waste biodegradation rate in a food waste biodigester with the synergistic action of hydrolase-producing *Bacillus paralicheniformis* GRA2 and *Bacillus velezensis* TAP5 co-culture inoculation. *Saudi Journal of Biological Sciences*, *28*(5), 3001–3012. https://doi.org/10.1016/J.SJBS.2021.02.041

Sanders, E. R. (2012). Aseptic laboratory techniques: Plating methods. *Journal of Visualized Experiments*, *63*, e3064. https://doi.org/10.3791/3064

Sarkar, S., Pal, S., & Chanda, S. (2016). Optimization of a vegetable waste composting process with a significant thermophilic phase. *Procedia Environmental Sciences*, *35*, 435–440. https://doi.org/10.1016/j.proenv.2016.07.026

Sethi, S., Datta, A., Lal Gupta, B., & Gupta, S. (2013). Optimization of cellulase production from bacteria isolated from soil. *International Scholarly Research Notices Biotechnology*, *2013*, 985685. https://doi.org/10.5402/2013/985685

Siles-Castellano, A. B., López, M. J., López-González, J. A., Suárez-Estrella, F., Jurado, M. M., Estrella-González, M. J., & Moreno, J. (2020). Comparative analysis of phytotoxicity and compost quality in industrial composting facilities processing different organic wastes. *Journal of Cleaner Production*, *252*, 119820. https://doi.org/10.1016/j.jclepro.2019.119820

Simair, A. A., Khushk, I., Qureshi, A. S., Bhutto, M. A., Chaudhry, H. A., Ansari, K. A., & Lu, C. (2017). Amylase production from thermophilic *Bacillus* sp. BCC 021-50 isolated from a marine environment. *Fermentation*, *3*(2), 25. https://doi.org/10.3390/FERMENTATION3020025

Singh, R., Kesherwani, B., Madhumita Sri, R. M., & Ravichandran, S. (2021). Sustainable management of food wastes using effective microorganisms compost. *Journal of Management Research and Analysis*, *8*(3), 127–130. https://doi.org/10.18231/j.jmra.2021.026

Storino, F., Arizmendiarrieta, J. S., Irigoyen, I., Muro, J., & Aparicio-Tejo, P. M. (2016). Meat waste as feedstock for home composting: Effects on the process and quality of compost. *Waste Management*, *56*, 53–62. https://doi.org/10.1016/J.WASMAN.2016.07.004

Tallapragada, P., Rayavarapu, B., Rao, P. P., Ranganath, N. N., & Veerabhadrappa, P. P. (2018). Screening of potential probiotic lactic acid bacteria and production of amylase and its partial purification. *Journal of Genetic Engineering and Biotechnology*, *16*(2), 357–362. https://doi.org/10.1016/j.jgeb.2018.03.005

Tsuchida, O., Yamagata, Y., Ishizuka, T., Arai, T., Yamada, J.-I., Takeuchi, M., & Ichishima, E. (1986). An alkaline proteinase of an alkalophilic *Bacillus* sp. *Current Microbiology*, *14*, 7–12. https://doi.org/10.1007/BF01568094

Van Dyk, J. S., Gama, R., Morrison, D., Swart, S., & Pletschke, B. I. (2013). Food processing waste: Problems, current management and prospects for utilisation of the lignocellulose component through enzyme synergistic degradation. *Renewable and Sustainable Energy Reviews*, *26*, 521–531. https://doi.org/10.1016/J.RSER.2013.06.016

Varma, V. S., Das, S., Sastri, C. V., & Kalamdhad, A. S. (2017). Microbial degradation of lignocellulosic fractions during drum composting of mixed organic waste. *Sustainable Environment Research*, *27*(6), 265–272. https://doi.org/10.1016/j.serj.2017.05.004

Wan, L., Wang, X., Cong, C., Li, J., Xu, Y., Li, X., Hou, F., Wu, Y., & Wang, L. (2020). Effect of inoculating microorganisms in chicken manure composting with maize straw. *Bioresource Technology*, *301*, 122730. https://doi.org/10.1016/J.BIORTECH.2019.122730

Wang, G., Yang, Y., Kong, Y., Ma, R., Yuan, J., & Li, G. (2022). Key factors affecting seed germination in phytotoxicity tests during sheep manure composting with carbon additives. *Journal of Hazardous Materials*, *421*, 126809. https://doi.org/10.1016/j.jhazmat.2021.126809

Wei, Y., Li, J., Shi, D., Liu, G., Zhao, Y., & Shimaoka, T. (2017). Environmental challenges impeding the composting of biodegradable municipal solid waste: A critical review. *Resources, Conservation and Recycling*, *122*, 51–65. https://doi.org/10.1016/J.RESCONREC.2017.01.024

Winkler, U. K., & Stuckmann, M. (1979). Glycogen, hyaluronate, and some other polysaccharides greatly enhance the formation of exolipase by *Serratia marcescens*. *Journal of Bacteriology*, *138*(3), 663–670. https://doi.org/10.1128/jb.138.3.663-670.1979

Wu, J., Zhao, Y., Zhao, W., Yang, T., Zhang, X., Xie, X., Cui, H., & Wei, Z. (2017). Effect of precursors combined with bacteria communities on the formation of humic substances during different materials composting. *Bioresource Technology*, *226*, 191–199. https://doi.org/10.1016/J.BIORTECH.2016.12.031

Yadav, V., Prakash, S., Srivastava, S., Verma, P. C., Gupta, V., Basu, V., & Rawat, A. K. (2009). Identification of *Comamonas* species using 16S rRNA gene sequence. *Bioinformation*, *3*(9), 381–383. https://doi.org/10.6026/97320630003381

Yafetto, L. (2022). Application of solid-state fermentation by microbial biotechnology for bioprocessing of agro-industrial wastes from 1970 to 2020: A review and bibliometric analysis. *Heliyon*, *8*(3), e09173. https://doi.org/10.1016/j.heliyon.2022.e09173

Yong, Z. J., Bashir, M. J. K., Ng, C. A., Sethupathi, S., Lim, J. W., & Show, P. L. (2019). Sustainable waste-to-energy development in Malaysia: Appraisal of environmental, financial, and public issues related with energy recovery from municipal solid waste. *Processes*, *7*(10), 676. https://doi.org/10.3390/PR7100676

Zainu, Z. A., & Songip, A. R. (2017). Policies, challenges and strategies for municipal waste management in Malaysia. *Journal of Science, Technology and Innovation Policy*, *3*(1), 10-14. https://doi.org/10.11113/jostip.v3n1.18

Zhang, L., & Sun, X. (2017). Addition of seaweed and bentonite accelerates the two-stage composting of green waste. *Bioresource Technology*, *243*, 154–162. https://doi.org/10.1016/J.BIORTECH.2017.06.099

Zhao, K., Xu, R., Zhang, Y., Tang, H., Zhou, C., Cao, A., Zhao, G., & Guo, H. (2017). Development of a novel compound microbial agent for degradation of kitchen waste. *Brazilian Journal of Microbiology*, *48*(3), 442–450. https://doi.org/10.1016/j.bjm.2016.12.011

Pertanika J. Sci. & Technol. 33 (6): 2521 - 2539 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

The Effects of Pre-Composting Subjected to Black Soldier Fly Larvae Digestion on Agri-Food Waste Maturation Indicators

Muhammad Heikal Ismail[1]*, Naweedullah Amin[2], Anie Yulistyorini[3], Ridwan Muhamad Rifai[3], Khursheed Alam Khan[4] and Luqman Chuah Abdullah[5]

[1]Department of Environment, Faculty of Forestry and Environment, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[2]Department of Zoology, Faculty of Biology, Kabul University, Kabul 1006, Afghanistan
[3]Department of Civil Engineering, Faculty of Engineering, Universitas Negeri Malang, Jalan Semarang No. 5, Malang 65145, Indonesia
[4]College of Horticulture, Mandsaur of R. V. S. Agriculture University, Gwalior 458001, India
[5]Department of Chemical and Environmental Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

ABSTRACT

The increasing demand for sustainable agricultural practices has led to the exploration of alternative composting methods, including the use of black soldier fly larvae (BSFL) for managing organic waste. This study compares the physical, chemical, nutrient, and nutrient loss properties of frass produced through two treatments at specified moisture contents of 65 and 80%, respectively (BSFL-65 and BSFL-80). The physical properties, including yield, water holding capacity, bulk density, and moisture content, were analyzed, revealing no significant differences between the treatments for yield, water holding capacity, and bulk density. However, moisture content was significantly higher in BSFL-80, indicating a distinct impact of treatment on water retention. In terms of chemical properties, the pH of BSFL-80 was significantly higher than that of BSFL-65, while electrical conductivity showed a significant difference at the borderline. Total dissolved solids were significantly higher in BSFL-80, whereas total volatile solids were significantly higher in BSFL-65. Nutrient content analysis revealed no significant differences in carbon (C), hydrogen (H), sulfur (S), nitrogen (N), phosphorus (P), potassium (K), and other elements, except for ammonia nitrogen, which was slightly higher in BSFL-80. Nutrient loss assessments showed no significant differences between the treatments for C, H, S, N, P, K, calcium, magnesium, copper, or zinc. These findings suggest that while certain chemical and physical properties, such as moisture content and pH, are significantly

ARTICLE INFO

Article history:
Received: 31 July 2024
Accepted: 17 April 2025
Published: 30 September 2025

DOI: https://doi.org/10.47836/pjst.33.6.06

E-mail addresses:
heikal@upm.edu.my (Muhammad Heikal Ismail)
sodes.amin123@gmail.com (Naweedullah Amin)
anie.yulistyorini.ft@um.ac.id (Anie Yulistyorini)
ridwan.muhamadrifai.ft@um.ac.id (Ridwan Muhamad Rifai)
khan_undp@yahoo.ca (Khursheed Alam Khan)
chuah@upm.edu.my (Luqman Chuah Abdullah)
* Corresponding author

© Universiti Putra Malaysia Press

influenced by the treatment type, the overall nutrient composition and nutrient loss were largely unaffected by the treatment variations. This study provides valuable insights into the comparative efficacy of BSFL-based frass treatments, offering potential implications for sustainable agricultural practices and waste management.

Keywords: BSFL, frass, nutrient, moisture-content, pre-composting

INTRODUCTION

Municipal solid waste (MSW) comprises 20 categories, including food waste, plastic, paper, and glass (Jalil, 2010). Among these, Food waste represents a significant global challenge, with approximately one-third of food produced for human consumption lost or wasted annually, leading to severe economic, social, and environmental repercussions (Mokrane et al., 2023). The economic cost of food waste reaches billions, impacting households and the entire supply chain, while socially, this waste starkly contrasts with the millions suffering from hunger (Raina & Bathla, 2024). Environmentally, food waste contributes to greenhouse gas emissions and resource depletion, with avoidable consumer food waste (ACFW) alone responsible for substantial emissions and land use (Coudard et al., 2024). Effective management strategies, including consumer education and innovative practices like anaerobic digestion and composting, are essential for mitigating these impacts (Trivedi et al., 2023). As awareness grows, addressing the multifaceted drivers of food waste—from cultural to economic factors—becomes crucial for global sustainability and food security (Raina & Bathla, 2024).

Composting food waste using BSFL has emerged as an effective method for waste reduction and organic fertilizer production. Studies indicate that BSFL can significantly degrade organic matter, with one study reporting a substrate reduction of 87.95% when 2000 larvae were applied to 20 kg of kitchen waste over 22 days (Mohod et al., 2024). The larvae consume food waste efficiently and enhance the resulting compost's nutritional profile, which meets agricultural standards for essential nutrients like N, P, and K (Putri et al., 2024). Furthermore, optimal conditions such as larval density and environmental factors can improve waste reduction rates, with some studies achieving reductions of up to 84.5% in organic waste (Amin et al., 2024). Overall, utilizing BSFL for composting presents a sustainable alternative to traditional waste management practices, contributing to both waste diversions from landfills and the production of high-quality organic fertilizers (Hariri et al., 2024).

Recent studies have also shown the promise of BSFL frass, the by-product of larvae, as a biofertilizer. For instance, basil plants grown with BSFL frass showed substantial growth when compared to those grown with synthetic fertilizers, although some deficiencies were noted, such as lower levels of manganese (Mn), calcium (Ca), and zinc (Zn) (Romano et al., 2024). Additionally, BSFL frass has been demonstrated to provide adequate nutrients

for lettuce growth, reducing the need for additional fertilizer applications after the initial growth cycle (Chiam et al., 2021). Furthermore, research has highlighted that BSFL frass is a suitable soil amendment, promoting plant growth while maintaining environmental safety, with no heavy metal contamination or pathogens found in the frass (Chiam et al., 2021). In a comparative study, the application of BSFL frass to Pakchoi resulted in greater biomass production than chemical fertilizers, supporting its potential as a viable alternative to synthetic fertilizers in sustainable agriculture (Agustiyani et al., 2021). These findings reinforce the potential of BSFL frass as a nutrient-rich and environmentally safe amendment, enhancing the role of BSFL in organic waste management.

Moisture content in food waste significantly influences its management and processing, with varying optimal levels identified across different studies. For instance, raw food waste typically contains around 40% moisture, which can be effectively treated using self-sustaining smoldering techniques, achieving over 90% mass destruction (Song et al., 2022). In composting, a moisture content of approximately 60% is optimal for maintaining peak temperatures and minimizing odor emissions (Rose et al., 2019). Conversely, studies indicate that moisture levels below 40% are preferable to mitigate odor production during decomposition (Qamaruz-Zaman et al., 2019). Pre-treatment methods, such as hot water washing, can reduce the moisture content from 76.7 to 48.9%, enhancing drying efficiency and environmental sustainability (Aziz & Qaisari, 2024).

Moisture content is a critical factor influencing the growth and survival of BSFL, *Hermetia illucens*, as it directly affects their performance on various substrates. Research indicates that optimal moisture levels enhance larval growth, with studies showing that larvae reared at 70% moisture exhibited superior growth metrics compared to those at lower levels (Okpoko et al., 2024). Additionally, moisture influences both larval survival and pupation rates, with higher moisture levels significantly reducing prepupal mortality and promoting deeper pupation (Liu et al., 2023). Furthermore, the water-release properties of substrates play a vital role, as substrates with appropriate moisture content correlate with increased larval survival (Frooninckx et al., 2024). The moisture content of food waste can also affect the efficiency of residue separation from the insect biomass, which is an important consideration for waste management and bioconversion technologies. Studies have shown that moisture contents of 70 and 75% facilitate effective residue separation through sieving, while higher moisture contents pose challenges to this process (Cheng et al., 2017).

Moreover, the specific moisture content of substrates has been linked to the growth and metabolic performance of BSFL. A study examining substrate moisture content from 45 to 85% found that larvae developed best at moisture contents ranging from 45 to 75%, with the highest growth rate observed at 45% moisture content. The performance of BSFL was found to be indirectly influenced by substrate moisture due to microbial processes that co-occur with larval development (Bekker et al., 2021). Overall, maintaining adequate

moisture levels is essential for maximizing the efficiency of BSFL in waste conversion processes, thereby supporting their potential as a sustainable protein source.

This study evaluates the feasibility and effectiveness of using BSFL as a sustainable approach for processing food waste, mainly fruits and vegetables, into nutrient-dense compost. By testing different moisture levels (65 and 80%, respectively), this study seeks to identify optimal conditions that enhance the nutrient content and quality of the resulting compost. Through a comprehensive analysis of physicochemical and nutritional properties, this research aims to establish frass produced by BSFL as a viable alternative for food waste management, thereby supporting agricultural productivity and environmental sustainability.

MATERIALS AND METHODS

Source of Materials and Processing

The food waste for this experiment was collected from HERO and BIG supermarkets in Sri Serdang, Selangor (Malaysia), and the local market. The food waste composition comprised spoiled vegetables and fruits, including cabbage, cucumber, tomato, banana, and papaya, at a ratio of 1:1. A total of 20 kg of food waste was collected. To ensure uniformity in the frass process, the size of the food waste was cut within the range of 2 to 3 cm.

The food waste was thermally pre-treated using an oven dryer at 70°C to achieve a moisture content of 60%. After the pre-treatment, the food waste was divided into three samples, each weighing 625 g, and mixed with garden soil in a ratio of 5:3 (food waste: garden soil). Once combined with the garden soil, the samples were poured into separate composting boxes sized 30 cm in length, 10 cm in height, and 12 cm in width to accommodate the mixture effectively. To initiate the pre-composting phase, water was added to each sample to achieve the specified moisture contents of 65 and 80%, respectively. The moisture content of the samples was monitored every two days using a soil moisture detector, and adjustments were made by adding water as needed to maintain the desired levels. Additionally, each composting box was covered with newspapers to prevent insect access, such as flies.

BSFL Digestion

Live BSFL was purchased online (Lazada, Malaysia) under the Biovae shop. The larval density was 1 – 2.5 larvae/g, and the weight was 0.2 ± 0.1 g. After the pre-composting phase, 20 g of BSFL was added to each composting box. To prevent contact with insects, all samples were covered with newspaper, ensuring the BSFL could decompose the food waste without interference. The composting boxes were kept in a low-light-intensity area for 15 days. The frass was harvested using a 2 mm sieve tray after the composting period. The frass containing the organisms was filled approximately one-third of the way in the sieve tray. The tray was gently shaken from side to side, similar to sifting flour. As a result, the frass

fell through the sieve. This sifting method was repeated until all samples were processed, ensuring the frass had a consistent texture and reducing its particle size from coarse to fine.

Physical Properties of Frass

The yield of frass produced was calculated by measuring the weight, which was then divided by the amount of food waste used. Water holding capacity of frass was measured using the following equation.

$$WHC = \frac{(Ws - Wi) + MC \times Wi}{(1 - MC) \times Wi}$$

Where, WHC = Water holding capacity (g water per g dry compost); W_i = Initial weight of frass (before filtration); W_s = Final weight of frass (after filtration); MC = Moisture content of frass

The bulk density of the frass and soil was measured by filling 150 g of frass into a 10 L container. The compost was then compressed until no large void spaces remained, and the volume occupied was measured. The bulk density was calculated by dividing the weight of the frass by the volume occupied (kg/m³).

Chemical Properties of Frass

A total of 20 g of the frass sample was weighed and placed in a 100 ml beaker. Subsequently, 20 ml of distilled water was added to the beaker, and the mixture was stirred continuously for 15 min. The beaker was then covered and allowed to stand undisturbed for 30 min. After the settling period, a pH meter was carefully inserted into the liquid region of the sample solution to measure and record the pH value. Electrical conductivity (EC) was measured using a portable EC detector (SBPHM001, OEM, China). Total dissolved solids (TDS) were measured using a portable TDS meter (TDS-3, China).

The total volatile solids (TVS) content in the frass was determined based on the results of the moisture content test. Initially, the frass samples were heated to a temperature of 105°C, as per the moisture content test procedure. Following this, the samples were subjected to ashing in a furnace at a temperature of 550°C for a duration of 4 hr. After the ashing process, the remaining ash was weighed, and this mass was used to represent the TVS content in the frass (Vijayan et al., 2024).

Nutrient Content

The elements in frass were analyzed using an auto analyzer and an atomic absorption spectrophotometer (AAS) (AAnalyst200, PerkinElmer, USA). The macro and micro-elements measured included C, H, N, K, P, Ca, magnesium (Mg), Zn, copper (Cu), and S.

For the determination of C, H, N, and S concentrations, solid samples of frass were sent to the Material Characterization Laboratory, Faculty of Engineering, Universiti Putra Malaysia (UPM). To measure the concentrations of P, K, Ca, Mg, Cu, and Zn, the samples were analyzed at Analytical Laboratory 2, Faculty of Agriculture, UPM, using an AAS.

The samples were prepared in liquid form using the dilute double acid (DDA) method. Specifically, 5.0 g of each sample was weighed and placed into a Falcon tube, followed by the addition of 25 ml of DDA extracting reagent. The DDA extracting reagent was prepared in the laboratory by mixing hydrochloric acid (HCl, 37%, Merck, Germany) and sulfuric acid (H_2SO_4, 98%, Sigma-Aldrich, USA) to obtain final concentrations of 0.05 N HCl and 0.025 N H_2SO_4, respectively. The falcon tube was shaken for 15 min using an end-to-end shaker at 180 rpm. Subsequently, the solution was filtered through filter paper, and the filtrate was collected in a 50 ml plastic vial. The filtrate was then sent for analysis using an AAS.

The calculation for ammoniacal nitrogen was as follows:

$$NH4 - N = \frac{Vs \times M \times 0.014 \times 100}{W}$$

Where, Vs = Volume of sulfuric acid used for sample titration; M = Molarity of sulfuric acid; W = Weight of dry soil sample

Nutrient Losses

A sieve tray was prepared for the extraction of nutrients from frass. To prevent the loss of compost material during the process, a layer of muslin cloth was securely placed inside the sieve tray. The muslin cloth acted as a filter, allowing water to pass through while retaining the solid compost particles. A uniform layer of frass, approximately 2.5 cm thick, was evenly spread over the muslin cloth in the sieve tray. Following this, 500 ml of distilled water was carefully poured onto the frass layer. The distilled water was used to ensure that no external contaminants interfered with the nutrient extraction process. As the distilled water percolated through the frass layer, it carried dissolved nutrients with it. The excess water, along with the leached nutrients, passed through the muslin cloth and sieve tray. This leachate was collected in a clean, empty container positioned beneath the sieve tray to ensure no loss of the extracted solution. The collected leachate was then subjected to nutrient analysis to determine the concentration and composition of the dissolved nutrients.

Statistical Analysis

All statistical analyses were performed using Python (3.13.0). Descriptive statistics were computed for all measured parameters. The normality of data distribution was assessed using the Shapiro-Wilk test. An independent samples *t*-test was conducted to compare the

means of variables between the two treatments (BSFL-1-65 and BSFL-1-80). A significance level of $p < 0.05$ was considered statistically significant.

RESULTS

Physical Properties

The physical properties of two treatments, BSFL-1-65 and BSFL-1-80, were assessed based on four factors: yield (%), water holding capacity (g/g), bulk density (kg/m³), and moisture content (%). For yield (%), the results were relatively similar across both treatments, with BSFL-1-65 showing a range from 48.76 to 50.88%, and BSFL-1-80 ranging from 49.15 to 51.88% as shown in Figure 1a. The *t*-test for yield revealed no significant difference between the two treatments (*t*-statistic = -0.6020, *p*-value = 0.5796). Water holding capacity also showed minimal variation between the two groups, with values ranging from 3.31 to 3.43 g/g for BSFL-1-65 and 3.35 to 3.43 g/g for BSFL-1-80 (Figure 1b). The t-test

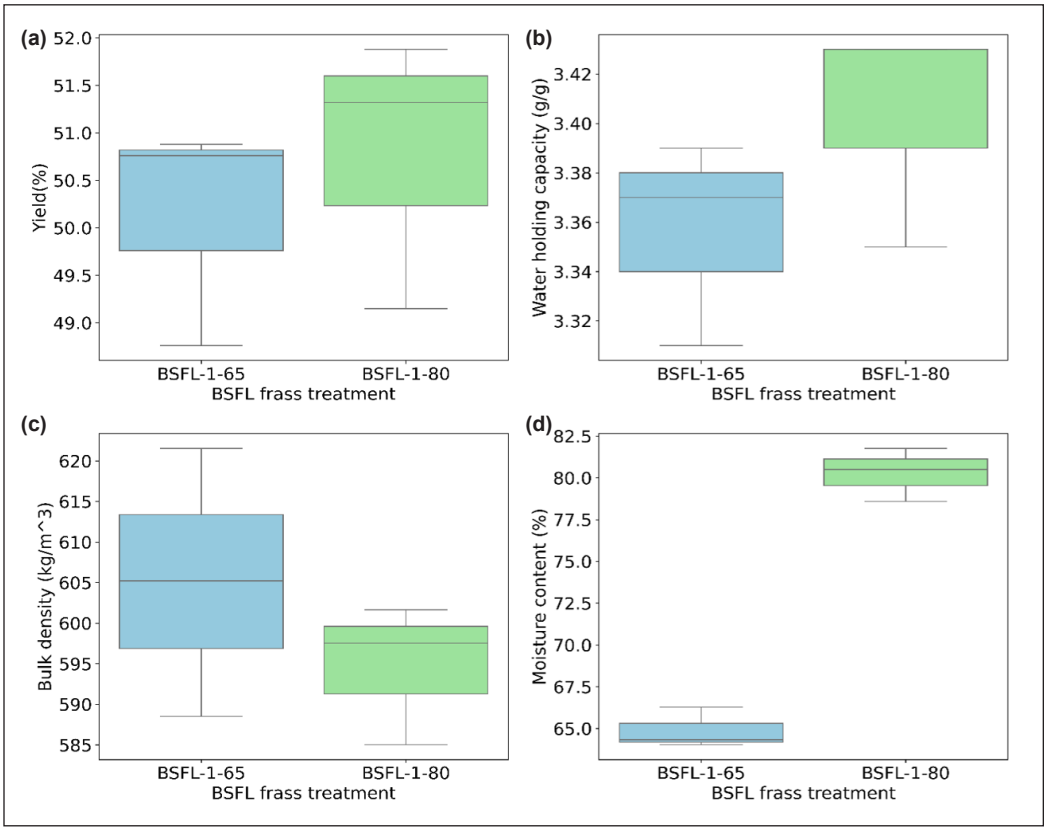


Figure 1. Physical properties of frass: comparison of yield percentage, water holding capacity, bulk density, and moisture content between BSFL-1-65 and BSFL-1-80 treatments

Note. BSFL = Black soldier fly larvae; No significant differences were observed across the physical properties ($p > 0.05$)

for this parameter also indicated no significant difference (t-statistic = -1.2999, p-value = 0.2635). Bulk density measurements were found to range from 588.51 to 621.53 kg/m³ for BSFL-1-65, and 585.02 to 601.65 kg/m³ for BSFL-1-80, with the t-test showing no significant difference between the treatments (t-statistic = 0.9611, p-value = 0.3909) (Figure 1c). However, significant differences were observed for moisture content (%), with BSFL-1-65 ranging from 64.03 to 66.29% and BSFL-1-80 ranging from 78.59 to 81.76% (Figure 1d). The t-test for moisture content yielded a highly significant result (t-statistic = -13.2477, p-value = 0.0002), indicating a marked difference between the two treatments. This suggests that moisture content is the only physical property significantly affected by the treatment type.

Chemical Properties

The chemical properties of two treatments, BSFL-1-65 and BSFL-1-80, were assessed based on pH, EC, TDS, and TVS. For pH, the results were quite similar across both treatments, with BSFL-1-65 ranging from 6.68 to 6.70, and BSFL-1-80 from 6.72 to 6.75 (Figure 2a). The t-test revealed a statistically significant difference in pH between the two treatments

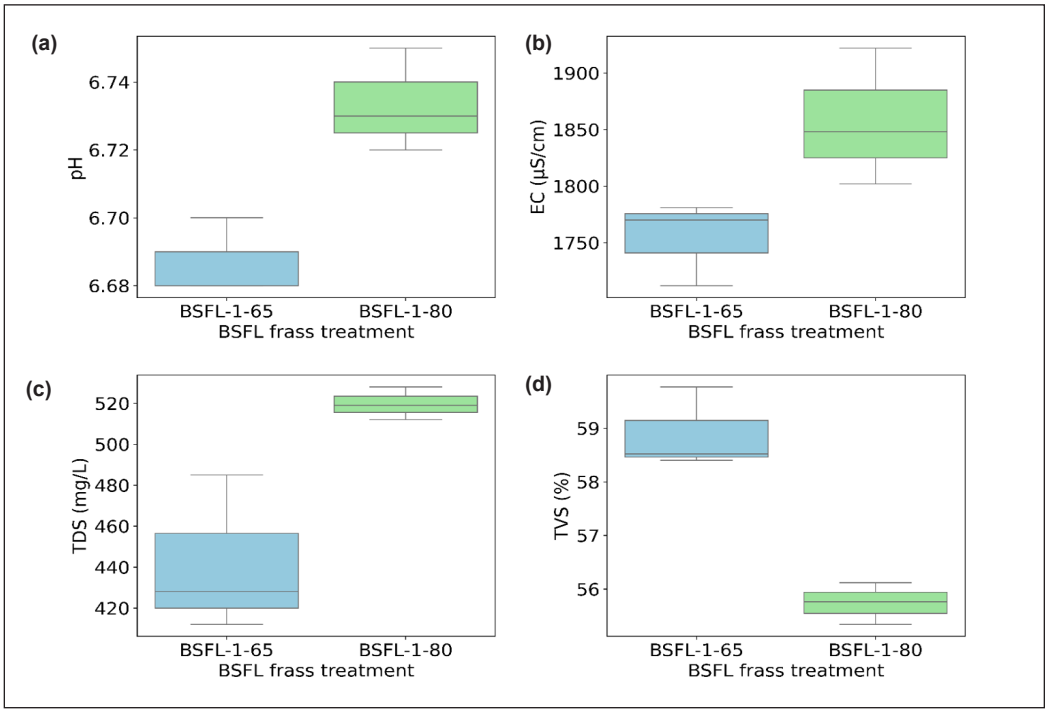


Figure 2. Chemical properties of frass: comparison of pH, electrical conductivity (EC), total dissolved solids (TDS), and total volatile solids (TVS) between BSFL-1-65 and BSFL-1-80 treatments
Note. BSFL = Black soldier fly larvae; Significant differences were observed in pH and TDS, with no significant differences in EC and TVS ($p < 0.05$)

(*t*-statistic = -4.2212, *p*-value = 0.0135), suggesting that the treatment notably influenced pH. EC values for BSFL-1-65 ranged from 1712 to 1781 µS/cm, while BSFL-1-80 ranged from 1802 to 1922 µS/cm (Figure 2b). The *t*-test for EC showed a borderline significance (*t*-statistic = -2.5130, *p*-value = 0.0658), indicating that while the difference between treatments is noticeable, it is not statistically strong at the conventional significance level. TDS values followed a similar trend, with BSFL-1-65 ranging from 412 to 485 mg/L and BSFL-1-80 ranging from 512 to 528 mg/L (Figure 2c). The *t*-test for TDS revealed a significant difference (*t*-statistic = -3.4464, *p*-value = 0.0261), indicating that TDS levels were significantly higher in BSFL-1-80. Lastly, TVS values ranged from 56.12 to 59.77% for BSFL-1-65 and from 55.34 to 55.76% for BSFL-1-80 (Figure 2d). The *t*-test for TVS showed a significant result (*t*-statistic = 6.4079, *p*-value = 0.0030), with BSFL-1-65 showing notably higher TVS values compared to BSFL-1-80. This suggests that TVS is the most significantly affected chemical property by the treatment type.

Nutrient Content

The nutrient content of frass from the BSFL-1-65 and BSFL-1-80 treatments was assessed across several parameters, including C, H, S, N, P, K, ammonia nitrogen (NH_3-N), Ca, Mg, Cu, Zn, carbon-to-nitrogen ratio (C/N), and nitrogen-to-phosphorus-to-potassium ratio (NPK). The C content for BSFL-1-65 ranged from 35.76 to 37.49%, while BSFL-1-80 ranged from 35.34 to 36.34%, with no significant difference between the two treatments (*t*-statistic = 1.1073, *p*-value = 0.3303) (Figure 3a). Similarly, H content varied from 2.32 to 2.66% in BSFL-1-65 and from 2.41 to 2.76% in BSFL-1-80, but this difference was not statistically significant (*t*-statistic = -0.4246, *p*-value = 0.6929) (Figure 3b). S levels ranged from 0.16 to 0.19% for BSFL-1-65 and 0.15 to 0.18% for BSFL-1-80, with no significant difference found (*t*-statistic = 1.2500, *p*-value = 0.2794) (Figure 3c). N concentrations were slightly higher in BSFL-1-65, ranging from 1.28 to 2.09%, compared to BSFL-1-80, which ranged from 1.31 to 2.12%, but the difference was not statistically significant (*t*-statistic = -0.1802, *p*-value = 0.8658) (Figure 3d). For P, values ranged from 1073.89 to 1739.60 mg/L in BSFL-1-65 and 1008.66 to 1835.84 mg/L in BSFL-1-80, with no significant differences observed (*t*-statistic = -0.2255, *p*-value = 0.8326) (Figure 3e). K concentrations ranged from 958.76 to 1739.60 mg/L in BSFL-1-65 and 1008.66 to 1835.84 mg/L in BSFL-1-80, also showing no significant difference (*t*-statistic = -0.3273, *p*-value = 0.7599) (Figure 3f). NH_3-N levels were slightly higher in BSFL-1-80, ranging from 49.91 to 65.29 mg/L, compared to BSFL-1-65 (52.47 to 64.10 mg/L), though the difference was not statistically significant (*t*-statistic = 0.0933, *p*-value = 0.9302) (Figure 3g). Ca content ranged from 3.29 to 6.88 mg/L in BSFL-1-65 and 5.95 to 7.18 mg/L in BSFL-1-80, with no significant difference (*t*-statistic = -1.0508, *p*-value = 0.3527) (Figure 3h). Mg levels in BSFL-1-65 ranged from 5.55 to 9.44 mg/L, while in

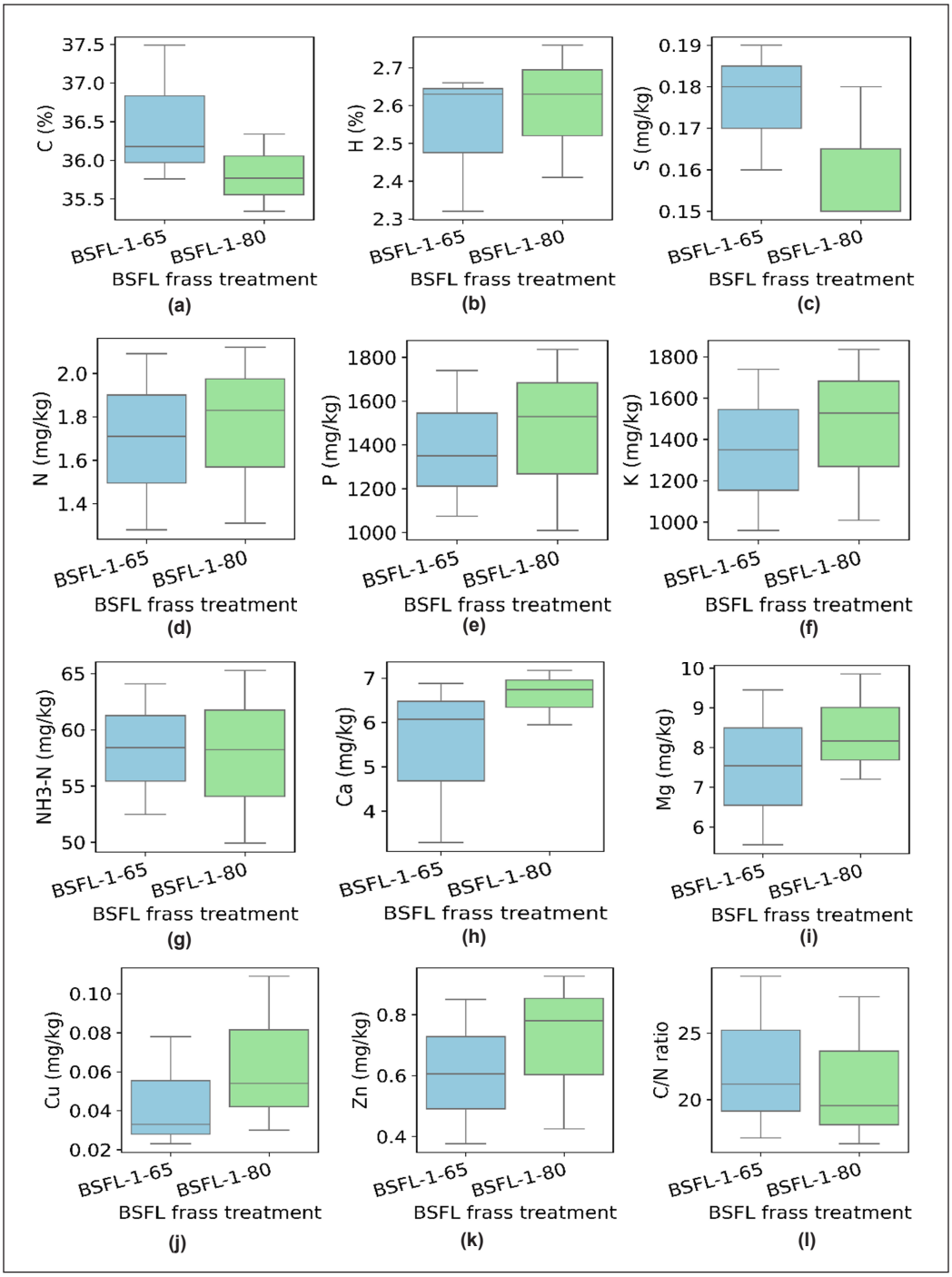


Figure 3. Nutrient content of frass: comparison of carbon (C), hydrogen (H), sulfur (S), nitrogen (N), phosphorus (P), potassium (K), ammonia nitrogen (NH_3-N), calcium (Ca), magnesium (Mg), copper (Cu), zinc (Zn), and carbon-to-nitrogen (C/N) ratio between BSFL-1-65 and BSFL-1-80 treatments
Note. BSFL = Black soldier fly larvae; No significant differences were found in the nutrient content ($p > 0.05$)

BSFL-1-80, they ranged from 7.20 to 9.85 mg/L, with no significant difference (*t*-statistic = -0.6564, *p*-value = 0.5474) (Figure 3i). Cu content was slightly higher in BSFL-1-80, ranging from 0.03 to 0.11 mg/L, compared to 0.023 to 0.078 mg/L in BSFL-1-65, though no significant difference was observed (*t*-statistic = -0.6815, *p*-value = 0.5330) (Figure 3j). Zn levels ranged from 0.376 to 0.85 mg/L in BSFL-1-65 and from 0.425 to 0.93 mg/L in BSFL-1-80, with no statistically significant difference (*t*-statistic = -0.4931, *p*-value = 0.6478) (Figure 3k). Finally, the C/N ratio ranged from 17.11 to 29.29 in BSFL-1-65 and from 16.67 to 27.74 in BSFL-1-80, with no significant difference (*t*-statistic = 0.2459, *p*-value = 0.8179) (Figure 3l).

The NPK ratio for both treatments remained consistent around 19:1:1 for BSFL-1-65 and 21:1:1 for BSFL-1-80, showing no significant difference in their composition (Table 1).

Table 1
NPK ratio of frass from BSFL-1-65 and BSFL-1-80 treatments

Treatment	NPK ratio
BSFL-65	19.2:1:1
	20.9:1:1
	20.4:1:1
BSFL-80	21.6:1:1
	20:1:1
	20.1:1:1

Note. NPK = Nitrogen-to-phosphorus-to-potassium ratio; BSFL = Black soldier fly larvae

Nutrient Loss

The nutrient loss content of frass from two treatments, BSFL-1-65 and BSFL-1-80, was evaluated across various parameters, including C, H, S, N, P, K, NH_3-N, Ca, Mg, Cu, and Zn. The results demonstrated no significant differences between the two treatments for any of the evaluated nutrients (Figure 4). Specifically, C loss ranged from 0.156 to 0.178% for BSFL-1-65 and 0.157 to 0.172% for BSFL-1-80 (t = 0.1541, p = 0.8860), while H loss ranged from 0.062 to 0.071% for BSFL-1-65 and 0.061 to 0.069% for BSFL-1-80 (t = 0.2683, p = 0.8021). S loss was 0.064 to 0.072% for BSFL-1-65 and 0.061 to 0.071% for BSFL-1-80 (t = -0.1250, p = 0.9080), and N loss ranged from 2.777 to 4.46% for BSFL-1-65 and 2.639 to 5.133% for BSFL-1-80 (t = -0.1480, p = 0.8905). For P, the loss was 1.08 to 3.015% for BSFL-1-65 and 1.08 to 2.418% for BSFL-1-80 (t = 0.3699, p = 0.7322). K loss ranged from 0.89 to 1.76% for BSFL-1-65 and 0.759 to 1.539% for BSFL-1-80 (t = 0.3837, p = 0.7211). NH_3-N loss varied from 5.047 to 8.109 mg/L for BSFL-1-65 and 4.786 to 9.308 mg/L for BSFL-1-80 (t = -0.1411, p = 0.8956). Ca loss was 0.287 to 0.529 mg/L for BSFL-1-65 and 0.292 to 0.56 mg/L for BSFL-1-80 (t = 0.5449, p = 0.6173). Mg loss ranged from 0.0558 to 0.0930 mg/L for BSFL-1-65 and 0.0620 to 0.0910 mg/L for BSFL-1-80 (t = 0.0790, p = 0.9409). Cu loss varied from 0.023 to 0.078 mg/L for BSFL-1-65 and 0.03 to 0.11 mg/L for BSFL-1-80 (t = 0.2403, p = 0.8227). Finally, Zn loss ranged from 0.333 to 0.505 mg/L for BSFL-1-65 and 0.315 to 0.409 mg/L for BSFL-1-80 (t = 0.6761, p = 0.5474).

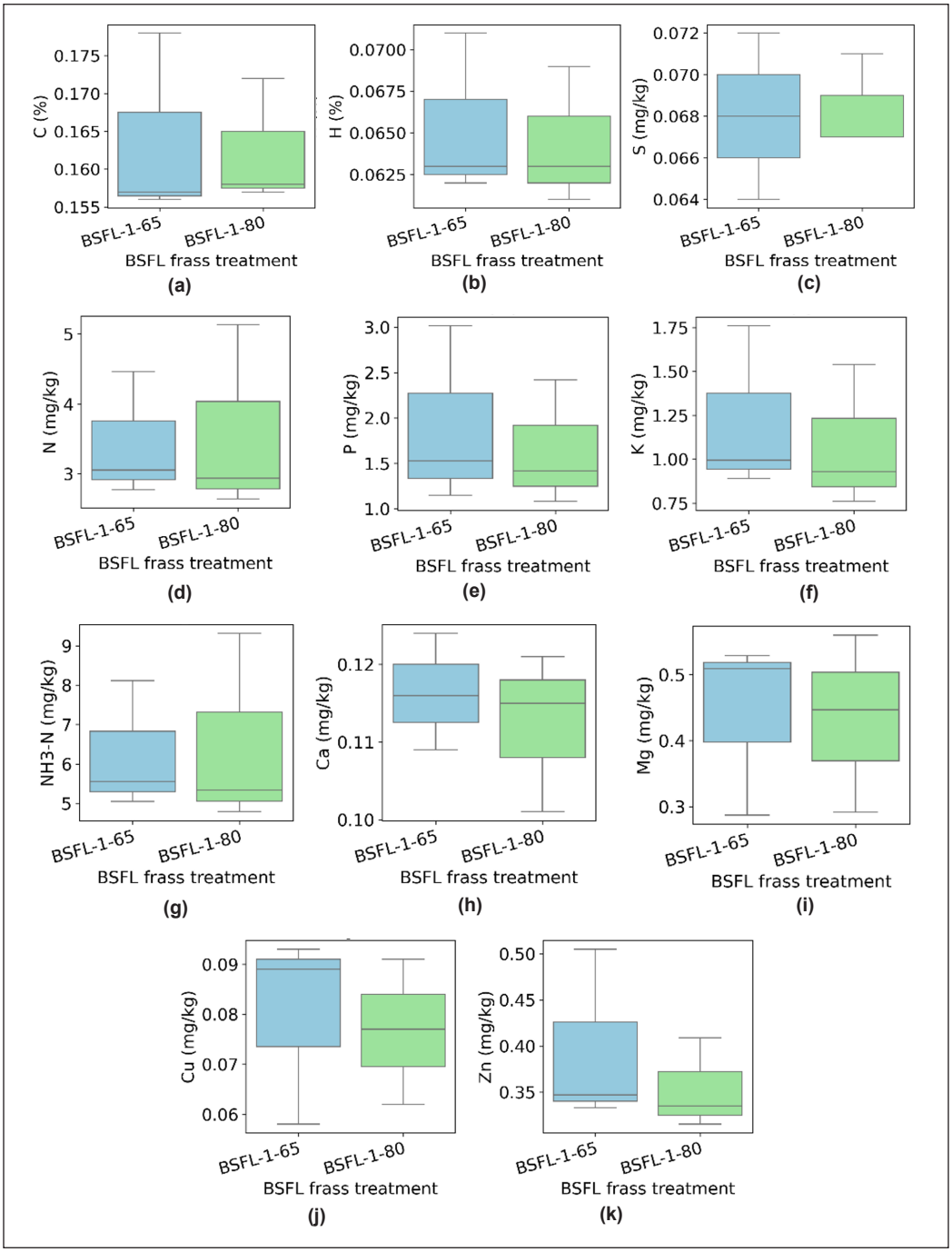


Figure 4. Nutrient loss content of frass: comparison of nutrient loss carbon (C), hydrogen (H), sulfur (S), nitrogen (N), phosphorus (P), potassium (K), ammonia nitrogen (NH_3-N), calcium (Ca), magnesium (Mg), copper (Cu), and zinc (Zn) between BSFL-1-65 and BSFL-1-80 treatments

Note. BSFL = Black soldier fly larvae; No significant differences were observed in nutrient loss across all parameters ($p > 0.05$)

DISCUSSION

In this study, the physical properties of two treatments, BSFL-1-65 and BSFL-1-80, were thoroughly evaluated based on yield (%), water holding capacity (g/g), bulk density (kg/m³), and moisture content (%), revealing significant differences in moisture content, while other parameters showed no significant variations between the treatments. Higher moisture content (80%) tends to enhance microbial activity, leading to improved degradation rates and higher yields, as microorganisms thrive in optimal moisture conditions, facilitating aerobic decomposition (Zavala & Funamizu, 2005). Conversely, at lower moisture levels (65%), the composting process may experience reduced biological activity, resulting in lower yields and diminished water holding capacity due to increased bulk density, which can hinder the compost's ability to retain moisture (Shyamala & Belagali, 2012; Zailani, 2018). Additionally, studies indicate that excessive moisture can lead to anaerobic conditions, negatively impacting compost quality and stability (Khater, 2012).

Varying moisture levels can influence the water holding capacity of frass, which is crucial for its application in irrigation systems. Research indicates that higher moisture content enhances the biodegradation rates of organic materials, leading to improved compost quality and water retention properties (Zavala & Funamizu, 2005). Compost application has been shown to increase soil water retention, with studies demonstrating that organic amendments can enhance water storage capacity, particularly in sandy soils (Le Guyader et al., 2024; Zemánek, 2014). For instance, a combination of biochar and compost resulted in a notable increase in water holding capacity, with optimal mixtures achieving substantial improvements in soil moisture retention (de Jesus Duarte et al., 2022). Furthermore, the effectiveness of compost in retaining moisture is influenced by its composition and the soil's textural characteristics, suggesting that higher moisture levels can further optimize the benefits of compost in irrigation systems (Negara et al., 2023; Zemánek, 2014).

Varying moisture content significantly influences water holding capacity and bulk density in frassing, with distinct effects observed at 65 and 80% moisture levels, respectively. At 65% moisture, composting processes tend to exhibit enhanced microbial activity, leading to accelerated organic matter degradation and improved humification, as evidenced by higher temperatures and microbial interactions (Sun et al., 2024). Conversely, at 80%, while microbial activity remains, the increased moisture can lead to reduced oxygen availability, potentially hindering aerobic processes and resulting in lower overall compost quality (Kim et al., 2015). Studies indicate that optimal moisture levels, typically around 70-75%, balance microbial efficiency and physical structure, promoting effective composting while minimizing greenhouse gas emissions (Zavala & Funamizu, 2005). Thus, maintaining an appropriate moisture content is crucial for optimizing composting outcomes, as it influences both water retention and bulk density.

For the chemical properties, both treatments, BSFL-1-65 and BSFL-1-80, were evaluated based on pH, EC, TDS, and TVS, revealing significant differences in pH, TDS, and TVS. At the same time, EC exhibited a borderline significant difference between the treatments. At higher moisture levels, studies indicate that pH tends to stabilize within a neutral range, while EC and TDS values generally increase, reflecting enhanced solubility of nutrients and minerals, which is crucial for microbial activity and nutrient availability (Ajaweed et al., 2022; Khater, 2012). For instance, vermicomposting at 78% moisture resulted in higher mineral N content and microfauna biomass, suggesting that moisture levels above 65% promote better nutrient cycling and microbial growth (Qin et al., 2021). Additionally, the total volatile solids decrease as moisture content increases, indicating more efficient decomposition processes at elevated moisture levels (Ojo et al., 2018). Research indicates that higher moisture content, such as 80%, tends to enhance microbial activity, leading to improved N mineralization and higher biomass of beneficial microfauna, which can positively influence compost quality (Qin et al., 2021). Additionally, moisture levels impact pH stability, with studies showing that compost remains alkaline throughout the process, which is crucial for nutrient availability (Ameen et al., 2016). Furthermore, moisture content affects the degradation rates of organic matter, with optimal levels facilitating better breakdown of total organic carbon and volatile solids, thereby enhancing compost maturity (Zailani, 2018; Zavala & Funamizu, 2005).

Variations in moisture content significantly influence the nutrient composition of frass, with optimal levels enhancing microbial activity and nutrient availability. Studies indicate that moisture contents around 70 to 75% facilitate better composting outcomes, promoting higher temperatures and microbial interactions that enhance organic matter degradation and nutrient retention, particularly N forms like Kjeldahl-N and nitrate-N (Sun et al., 2024). For instance, a moisture content of 75% was found to optimize temperature increases and nutrient profiles, while lower moisture levels (below 65%) hindered aerobic decomposition and nutrient retention (Palsania et al., 2008; Zavala & Funamizu, 2005). Conversely, excessively high moisture levels (e.g., 80%) can lead to diminished microbial efficiency and increased greenhouse gas emissions, particularly nitrous oxide (N_2O) (Shi et al., 2012).

Moisture content significantly influences the NPK ratio, C/N ratio, and nutrient dynamics during composting processes. Higher moisture levels, such as 65%, have been shown to reduce gas emissions and nutrient losses, thereby enhancing nutrient retention, including N content, which increased by 33.3% during composting (Ghanney et al., 2021). The C/N ratio is also affected; for instance, chicken manure composting achieved a C/N ratio of 10, indicating optimal conditions for microbial activity. Additionally, initial moisture content impacts the degradation of organic matter, with studies indicating that moisture levels around 70% can accelerate the heating rate and enhance microbial interactions, leading to improved humification and nutrient availability (Sun et al., 2024). Furthermore,

moisture influences the grinding ratio of raw materials, which correlates negatively with C/N and total organic carbon, affecting overall compost quality (Wang et al., 2018).

The influence of composting moisture levels on N mineralization and ammonia volatilization is significant, with varying effects observed at 65 and 80% moisture content, respectively. At 65% moisture, composting demonstrated a 33.3% increase in total N content and reduced gaseous emissions, including ammonia, by 100% during the late mesophilic phase. Conversely, higher moisture levels (80%) were associated with increased ammonia volatilization, particularly when urea was applied, as evidenced by the highest ammonia (NH_3) fluxes recorded at 50% water-filled pore space (Castellano-Hinojosa et al., 2019). Additionally, moisture content impacts the kinetic and thermodynamic parameters of ammonia volatilization, with interactions between moisture and temperature significantly affecting reaction rates (Tao et al., 2017).

CONCLUSION

This study assessed the use of BSFL for frassing food waste at two different moisture levels (65 and 80%). The findings underscore the critical role of moisture content in influencing the compost's physical, chemical, and nutrient properties. While both moisture levels produced similar physical outcomes, the 80% treatment retained more moisture, highlighting the need for moisture control in optimizing composting processes. The 80% moisture level also showed higher nutrient solubility and microbial activity, whereas the 65% moisture treatment demonstrated more efficient organic matter decomposition. Both moisture treatments produced nutrient-rich frass, with the 65% treatment showing slightly higher retention of N and C, which are important for agricultural use. The study also found no significant nutrient loss between treatments, confirming that BSFL composting is effective in minimizing nutrient leaching. Overall, BSFL composting offers a sustainable alternative to traditional waste management practices, providing high-quality organic fertilizers and contributing to food waste diversion. In conclusion, optimizing moisture content in BSFL composting improves compost quality and nutrient recovery, with the 65% treatment being more efficient in organic matter decomposition and nutrient retention. These insights are useful for scaling up BSFL composting systems, especially in areas with food waste management challenges. Future research should focus on the long-term effects of BSFL compost on soil health and crop productivity to further establish this method as a sustainable solution for food waste management and environmental sustainability.

ACKNOWLEDGEMENTS

We want to thank Universiti Putra Malaysia for supporting this study under MATCHING GRANT/2023/9300493 and financing the publication under Dana Penerbitan Jurnal (9001103).

REFERENCES

Agustiyani, D., Agandi, R., Arinafril., Nugroho, A. A., & Antonius, S. (2021). The effect of application of compost and frass from Black Soldier Fly Larvae (*Hermetia illucens* L.) on growth of Pakchoi (*Brassica rapa* L.). In *IOP Conference Series: Earth and Environmental Science* (Vol. 762, No. 1, p. 012036). IOP Publishing. https://doi.org/10.1088/1755-1315/762/1/012036

Ajaweed, A. N., Hassan, F. M., & Hyder, N. H. (2022). Evaluation of physio-chemical characteristics of bio fertilizer produced from organic solid waste using composting bins. *Sustainability*, *14*(8), 4738. https://doi.org/10.3390/su14084738

Ameen, A., Ahmad, J., & Raza, S. (2016). Effect of pH and moisture content on composting of municipal solid waste. *International Journal of Scientific and Research Publications*, *6*(5), 35–37.

Amin, U. K., Lando, A. T., & Djamaluddin, I. (2024). Potential of black soldier fly larvae in reduction various types organic waste. *Ecological Engineering and Environmental Technology*, *25*(9), 190–201. https://doi.org/10.12912/27197050/190639

Aziz, K., & Qaisari, S. N. (2024). Hot water washing pre-treatment effects on hot air-drying process of food waste. *Journal for Research in Applied Sciences and Biotechnology*, *3*(1), 57–59. https://doi.org/10.55544/jrasb.3.1.9

Bekker, N. S., Heidelbach, S., Vestergaard, S. Z., Nielsen, M. E., Riisgaard-Jensen, M., Zeuner, E. J., Bahrndorff, S., & Eriksen, N. T. (2021). Impact of substrate moisture content on growth and metabolic performance of black soldier fly larvae. *Waste Management*, *127*, 73-79. https://doi.org/10.1016/j.wasman.2021.04.028

Castellano-Hinojosa, A., González-López, J., Vallejo, A., & Bedmar, E. J. (2019). Linking ammonia volatilization with moisture content and abundance of nitrification and denitrification genes in N-fertilized soils. In D. Zúñiga-Dávila, F. González-Andrés, & E. Ormeño-Orrillo (Eds.), *Microbial probiotics for agricultural systems* (pp. 29–43). Springer. https://doi.org/10.1007/978-3-030-17597-9_3

Cheng, J. Y. K., Chiu, S. L. H., & Lo, I. M. C. (2017). Effects of moisture content of food waste on residue separation, larval growth and larval survival in black soldier fly bioconversion. *Waste Management*, *67*, 315–323. https://doi.org/10.1016/j.wasman.2017.05.046

Chiam, Z., Lee, J. T. E., Tan, J. K. N., Song, S., Arora, S., Tong, Y. W., & Tan, H. T. W. (2021). Evaluating the potential of okara-derived black soldier fly larval frass as a soil amendment. *Journal of Environmental Management*, *286*, 112163. https://doi.org/10.1016/j.jenvman.2021.112163

Coudard, A., Sun, Z., Behrens, P., & Mogollón, J. M. (2024). The global environmental benefits of halving avoidable consumer food waste. *Environmental Science and Technology*, *58*(31), 13707–13716. https://doi.org/10.1021/acs.est.4c04140

de Jesus Duarte, S., Hubach, A., & Glaser, B. (2022). Soil water balance and wettability methods in soil treated with biochar and/or compost. *Carbon Research*, *1*, 31. https://doi.org/10.1007/s44246-022-00032-2

Frooninckx, L., Broeckx, L., Goossens, S., Wuyts, A., & Van Miert, S. (2024). Optimizing substrate moisture content for enhanced larval survival and growth performance in *Hermetia illucens*: Exploring novel approaches. *Discover Animals*, *1*, 7. https://doi.org/10.1007/s44338-024-00005-2

Ghanney, P., Qiu, H., Anning, D. K., Yang, H., Wang, Y., & Kugbe, J. X. (2021). Moisture-induced pattern of gases and physicochemical indices in corn straw and cow manure composting. *Applied Sciences*, *11*(18), 8493. https://doi.org/10.3390/app11188493

Hariri, M. N., Ishak, A. R., & Md Rashid, R. I. (2024). Effect of food waste on the growth performance, waste reduction efficiency and nutritional composition of black soldier fly (*Hermetia illucens* (L.), Diptera: Stratiomyidae) larvae. *Malaysian Applied Biology*, *53*(1), 137–146. https://doi.org/10.55230/mabjournal.v53i1.2713

Jalil, M. A. (2010). Sustainable development in Malaysia: A case study on household waste management. *Journal of Sustainable Development*, *3*(3), 91–102. https://doi.org/10.5539/jsd.v3n3p91

Khater, E.-S. G. (2012). Chemical and physical properties of compost. *Misr Journal of Agricultural Engineering*, *29*(4), 1567–1582. https://doi.org/10.21608/mjae.2012.101382

Kim, E., Lee, D.-H., Won, S., & Ahn, H. (2015). Evaluation of optimum moisture content for composting of beef manure and bedding material mixtures using oxygen uptake measurement. *Asian-Australasian Journal of Animal Sciences*, *29*(5), 753–758. https://doi.org/10.5713/ajas.15.0875

Le Guyader, E., Morvan, X., Miconnet, V., Marin, B., Moussa, M., Intrigliolo, D. S., Delgado-Iniesta, M. J., Girods, P., Fontana, S., Sbih, M., Boumaraf, B., Tirichine, A., Kavvadias, V., & Gommeaux, M. (2024). Influence of date palm-based biochar and compost on water retention properties of soils with different sand contents. *Forests*, *15*(2), 304. https://doi.org/10.3390/f15020304

Liu, Z., Morel, P. C. H., & Minor, M. A. (2023). Substrate and moisture content effects on pupation of the black soldier fly (Diptera: Stratiomyidae). *Journal of Insects as Food and Feed*, *9*(4), 415–426. https://doi.org/10.3920/JIFF2022.0057

Mohod, V. D., Bakane, P. H., Nage, S. M., Soneji, I. B., Jadhao, S. M., Undirwade, D. B., Satpute, N. S., & Zatale, N. D. (2024). *Biodegradation of kitchen waste into organic fertilizer using the black soldier fly (*Hermetia illucens*) larvae*. Research Square. https://doi.org/10.21203/rs.3.rs-4297462/v1

Mokrane, S., Buonocore, E., Capone, R., & Franzese, P. P. (2023). Exploring the global scientific literature on food waste and loss. *Sustainability*, *15*(6), 4757. https://doi.org/10.3390/su15064757

Negara, I. D. G. J., Salehudin, S., Yasa, I. W., Sulistiyono, H., Supriyadi, A., & Andriawan, L. D. A. (2023). Pengaruh durasi irigasi tetes terhadap reduksi lengas tanah pada lahan polybag [The effect of drip irrigation duration on soil moisture reduction in polybag land]. *Ganec Swara*, *17*(3), 1000.

Ojo, A. O., Taiwo, L. B., Adediran, J. A., Oyedele, A. O., Fademi, I., & Uthman, A. C. O. (2018). Physical, chemical and biological properties of an accelerated cassava based compost prepared using different ratios of cassava peels and poultry manure. *Communications in Soil Science and Plant Analysis*, *49*(14), 1774–1786. https://doi.org/10.1080/00103624.2018.1474914

Okpoko, O., Ebenebe, C. I., Okeke, J. J., Amobi, M. I., & Umeaniebue, A. C. (2024). Effect of substrate moisture on the growth performance and survivability of the black soldier fly (*Hermetia illucens* L.) larva. *International Journal of Research and Innovation in Applied Science*, *IX*(V), 1–8. https://doi.org/10.51584/IJRIAS.2024.905001

Palsania, J., Sharma, R., Srivastava, J. K., & Sharma, D. (2008). Effect of moisture content variation over kinetic reaction rate during vermicomposting process. *Applied Ecology and Environmental Research*, *6*(2), 49–61. https://doi.org/10.15666/aeer/0602_049061

Putri, D. A., Chadirin, Y., & Yuwono, A. S. (2024). Food waste management utilizing black soldier fly larvae. *Emerging Science Innovation*, *2*, 26–35. https://doi.org/10.46604/emsi.2023.12785

Qamaruz-Zaman, N., Abdul-Sukor, N. S., Ab.-Rahman, S. A., & Yaacof, N. (2019). The influence of moisture content on the production of odor from food waste using path analysis. *Environmental Science and Pollution Research*, *26*, 13658–13663. https://doi.org/10.1007/s11356-019-04808-9

Qin, J., Fu, X., Chen, X., & Cui, G. (2021). Changes in physicochemical properties and microfauna community during vermicomposting of municipal sludge under different moisture conditions. *Environmental Science and Pollution Research*, *28*, 31539–31548. https://doi.org/10.1007/s11356-021-12846-5

Raina, A., & Bathla, G. (2024). Harvesting solutions: An inquiry into food waste management for a sustainable future. In A. E. Jimenez Ruiz, S. Bhartiya, & V. Bhatt (Eds.), *Promoting sustainable gastronomy tourism and community development* (pp. 214–228). IGI Global. https://doi.org/10.4018/979-8-3693-1814-0.ch013

Romano, N., Pande, G. S. J., Datta, S. N., Sinha, A. K., & Islam, S. (2024). Basil (*Ocimum basilicum*) growth and biochemical attributes are influenced by various black soldier fly (*Hermetia illucens*) larvae frass types. *Journal of Insects as Food and Feed*, *10*(7), 1223–1233.

Rose, J., Jia Chi, L., Yi Lung, T., & Vithanawasam, C. K. (2019). Calibrated analytical model of low cost YL-69 hygrometer sensor for determining moisture content of fruit and vegetable waste. In *International UNIMAS STEM 12th Engineering Conference* (pp. 71–76). IEEE. https://doi.org/10.1109/EnCon.2019.8861253

Shi, L.-L., Shen, M.-X., Chang, Z.-Z., Wang, H.-H., Lu, C.-Y., Chen, F.-S., & Song, J. (2012). Effect of water content on composition of water hyacinth *Eichhornia crassipes* (Mart.) Solms residue and greenhouse gas emission. *Chinese Journal of Eco-Agriculture*, *20*(3), 337–342. https://doi.org/10.3724/SP.J.1011.2012.00337

Shyamala, D. C., & Belagali, S. L. (2012). Studies on variations in physico-chemical and biological characteristics at different maturity stages of municipal solid waste compost. *International Journal of Environmental Sciences*, *2*(4), 1984–1997.

Song, Z., He, T., Li, M., Wu, D., & You, F. (2022). Self-sustaining smoldering as a novel disposal approach for food waste with high moisture content. *Fuel Processing Technology*, *228*, 107144. https://doi.org/10.1016/j.fuproc.2021.107144

Sun, S., Guo, C., Wang, J., Ren, L., Qu, J., Guan, Q., Dou, N., Zhang, J., Chen, Q., Wang, Q., Wang, J., Li, J., Gao, Z., & Zhou, B. (2024). Effect of initial moisture content, resulting from different ratios of vegetable waste to maize straw, on compost was mediated by composting temperatures and microbial communities at low temperatures. *Chemosphere*, *357*, 141808. https://doi.org/10.1016/j.chemosphere.2024.141808

Tao, L., Xianghong, G., Juanjuan, M., Xihuan, S., Yang, F., & Hongyu, W. (2017). Kinetic and thermodynamic effects of moisture content and temperature on the ammonia volatilization of soil fertilized with urea. *International Journal of Agricultural and Biological Engineering*, *10*(6), 134–143. https://doi.org/10.25165/j.ijabe.20171006.3232

Trivedi, S., Bhoyar, V., Belgamwar, V., Wadher, K., A. Raut, N., & Dhoble, S. J. (2023). Practices of food waste management and its impact on environment. In N. A. Raut, D. M. Kokare, B. A. Bhanvase, K. R. Randive, & S. J. Dhoble (Eds.), *360-degree waste management* (Vol. 1, pp. 89–111). Elsevier. https://doi.org/10.1016/B978-0-323-90760-6.00001-1

Vijayan, V., Sama, K. C., Senger, A. S., Gopinath, A., Tyagi, M., & Saumya. (2024). *Standard operating procedure for testing wastewater, faecal sludge and biosolids*. Centre for Science and Environment. https://www.cseindia.org/standard-operating-procedure-for-testing-wastewater-faecal-sludge-and-biosolids-12076

Wang, L., Huang, Q., Wang, S., Li, G., & Yang, X. (2018). Relationships among raw materials, grinding ratios, and moisture content during the composting process. *Polish Journal of Environmental Studies*, *28*(1), 343–348. https://doi.org/10.15244/pjoes/81294

Zailani, S. N. (2018). *Moisture content influence on composting parameters and degradation kinetic models in an aerated closed system* [Doctoral thesis, Universiti Teknologi Malaysia]. Universiti Teknologi Malaysia Institutional Repository. http://eprints.utm.my/id/eprint/79192/

Zavala, M. A. L., & Funamizu, N. (2005). Effect of moisture content on the composting process in a biotoilet system. *Compost Science and Utilization*, *13*(3), 208–216. https://doi.org/10.1080/1065657X.2005.10702242

Zemánek, P. (2014). Evaluation of compost influence on soil water retention. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *59*(3), 227–232. https://doi.org/10.11118/actaun201159030227

Pertanika J. Sci. & Technol. 33 (6): 2541 - 2563 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

State-of-the-Art CNN Models for Plant Disease Classification: A Comparative Study

Tanko Daniel Salka[1], Marsyita Hanafi[1]*, Sharifah M. Syed Ahmad Abdul Rahman[1] and Dzarifah Mohamed Zulperi[2]

[1]Department of Computer and Communication System Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[2]Department of Plant Protection, Faculty of Agriculture, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

ABSTRACT

Plant diseases pose a significant threat to global food security, resulting in substantial agricultural losses. Traditional methods of diagnosing plant diseases rely on manual observation, which is time-consuming and prone to errors. Recent advancements in deep learning, particularly Convolutional Neural Networks (CNNs), offer a promising solution for automating and improving the accuracy of plant disease classification through image analysis. This study evaluates the performance of five state-of-the-art CNN architectures: VGG16, GoogleNet, EfficientNet B0, ResNet50, and DenseNet201 for plant leaf classification using the PlantVillage dataset, which comprises 54,305 images across 38 classes. Among the models, VGG16 achieved the highest accuracy of 93.75% and recall of 93.75%, with a low loss of 0.48, though it has a larger parameter size of 56.79 MB. GoogleNet followed closely with 88% accuracy, a high F1 score of 0.93, and a balanced size of 85.51 MB, despite a higher loss of 0.96. ResNet50 demonstrated strong performance with 84.36% accuracy and 0.35 loss, but was the most resource-intensive at 100.17 MB. EfficientNet B0, the smallest model at 18.62 MB, achieved 84.37% accuracy, whereas DenseNet201 underperformed, attaining only 69.32% accuracy and 1.06 loss, despite its moderate size of 28.01 MB. These findings highlight trade-offs between accuracy and computational efficiency, with VGG16 and GoogleNet excelling in precision, while EfficientNet B0 presents a compact alternative for resource-limited settings. This study provides valuable insights for selecting optimal CNN models for plant disease detection based on specific agricultural needs.

Keywords: Convolutional Neural Networks (CNN), deep learning model, plant disease classification, PlantVillage

ARTICLE INFO

Article history:
Received: 16 November 2024
Accepted: 29 April 2025
Published: 30 September 2025

DOI: https://doi.org/10.47836/pjst.33.6.07

E-mail addresses:
tankodaniel45@gmail.com (Tanko Daniel Salka)
marsyita@upm.edu.my (Marsyita Hanafi)
s_mumtazah@upm.edu.my (Sharifah M. Syed Ahmad Abdul Rahman)
dzarifah@upm.edu.my (Dzarifah Binti Mohamed Zulperi)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

The Food and Agriculture Organization (FAO) (2021) estimates that plant diseases contribute to approximately 1.3 billion tons of food waste annually, accounting for one-third of all food produced (Junaid & Gokce, 2024). The timely and accurate identification of these diseases is crucial for implementing effective control strategies, minimizing agricultural yield losses, and promoting sustainable farming methods. Traditionally, plant disease diagnosis has relied on manual inspections by skilled agronomists, a procedure that is both labor-intensive and prone to human error. The accuracy of these inspections is often compromised by factors such as inspector fatigue, limited expertise, and the subtlety of disease symptoms at early stages (Demilie, 2024). Moreover, monitoring large-scale agricultural fields poses significant challenges. Environmental factors, including variable lighting conditions, image quality, and resolution, further complicate disease detection (Gomez et al., 2024). Poor lighting or occlusions can distort disease symptoms, making accurate diagnosis difficult (Javidan et al., 2024). Consequently, traditional methods often fail to facilitate timely interventions, which are essential for mitigating crop losses.

Real-time disease detection in the field presents a significant challenge. In addition to effectively processing vast amounts of image data, a robust system must operate effectively under diverse field conditions, such as uneven lighting, varying weather patterns, and different crop growth stages. Addressing these complexities requires sophisticated algorithms that can distinguish between healthy and unhealthy crops in dynamic settings, ensuring timely and accurate diagnosis. Recent advancements in deep learning (DL), particularly CNNs, have revolutionized plant disease classification (L. C. Ngugi et al., 2021). These methods automate disease identification, offering a more accurate and efficient approach to diagnosing plant diseases through image analysis. DL is an advanced subset of machine learning (ML) that excels at handling complex problems involving diverse data types, such as videos and images.

ML, a subset of artificial intelligence (AI), enables systems to learn from data and improve performance without explicit programming. Its more advanced form, DL, employs multi-layered neural networks to process complex datasets, such as images and time-series data. DL has transformed various fields, including agriculture and healthcare. In agriculture, DL facilitates the automated detection of plant diseases, thereby enhancing crop yield and sustainability (L. C. Ngugi et al., 2021). Similarly, in healthcare, DL models have revolutionized disease diagnosis, drug discovery, and antimicrobial resistance management. For instance, Khan et al. (2023) reviewed AI's role in drug discovery, highlighting its ability to predict molecular interactions and combat resistant pathogens, while Alzubaidi et al. (2021) surveyed DL architectures, demonstrating their versatility in medical imaging and genomics. These cross-disciplinary successes underscore DL's potential to address complex classification tasks, including plant disease identification. By adapting pre-trained

CNN architectures, this research integrates cutting-edge DL techniques into agricultural applications, mirroring the progress seen in healthcare. DL techniques have gained significant attention due to their ability to deliver more accurate predictions than traditional ML algorithms. Recent efforts have aimed to enhance plant disease detection by refining deep-learning algorithms (Abouelmagd et al., 2024; Chowdhury et al., 2024; Perumal et al., 2024; Sofuoğlu & Birant, 2024). The increasing adoption of DL architectures for plant disease identification highlights the need for further research into novel deep-learning architectures. There is a growing demand for optimized models with fewer parameters, faster training times, and uncompromised performance. Many studies focus on evaluating individual models or using datasets with restricted variability, resulting in a lack of comprehensive comparative analyses of advanced CNN architectures under standardized conditions. In summary, the research problems that have inspired this study are:

- Limitations of traditional plant disease diagnosis, in which manual methods are labor-intensive, subjective, and prone to errors.
- Effective detection requires efficient data processing, adaptability to diverse field conditions, and robust algorithms for accurate classification in dynamic environments.

This research aims to bridge this gap by systematically assessing five prominent CNN models: VGG16, GoogleNet, EfficientNet B0, ResNet50, and DenseNet201, under various imaging conditions using the widely recognized PlantVillage dataset. The study evaluates model performance across multiple metrics, including accuracy, loss, recall, F1 score, precision, and parameter size, to determine DL architectures with fewer parameters and faster training times while maintaining high classification performance for plant leaf disease detection. Rather than proposing a new model, this study employs transfer learning with pre-trained CNN architectures to benchmark their effectiveness in plant leaf disease classification.

RELATED WORK

CNNs have significantly advanced plant disease detection; however, prior studies vary widely in scope, methodology, and practical applicability. This section synthesizes key works, critically evaluating their strengths and limitations to contextualize this study. While early efforts focused on accuracy (e.g., Le et al., 2020), recent research emphasizes efficiency and real-world deployment (Parez et al., 2025), revealing gaps in comprehensive multi-metric comparisons and architectural diversity. For example, Le et al. (2020) achieved 98.63% accuracy using a support vector machine (SVM) with a filtered local binary patterns (LBP) method with contour mask and coefficient k (k-FLBPCM) features, but its reliance on undistorted images limits robustness, making it less relevant to field conditions compared to CNN−based approaches. Bhagat et al. (2020) developed an optimized SVM-based method

for classifying plant leaves as healthy or unhealthy using Grid Search, improving accuracy from 80 to 84% while reducing computational costs. CNN-based models have shown superior performance. Ahmed et al. (2023) proposed a modified DenseNet201 architecture for grape leaf classification, evaluating the impact of layer freezing on fine-tuning. Using a dataset comprising 500 images across five categories, each containing 100 images, their DenseNet-30 model outperformed previous approaches with 98% accuracy. Similarly, Maeda-Gutiérrez et al. (2020) evaluated state-of-the-art architectures, including GoogleNet, AlexNet, ResNet18, ResNet50, and Inception V3, on the PlantVillage dataset featuring nine tomato diseases and a healthy class. GoogleNet excelled with an AUC of 99.72% and a sensitivity of 99.12%, demonstrating its effectiveness in detecting tomato diseases.

Other studies introduced novel modifications to existing architectures. Chen et al. (2020) developed MobileNet-Beta, an enhanced version of MobileNetV2 incorporating Classification Activation Map, achieving 99.85% accuracy on the PlantVillage dataset and 99.11% on an external dataset. Rinu and Manjula (2021) introduced a VGG16-based model for identifying 38 plant diseases, attaining a mean accuracy of 94.80% using images from the PlantVillage dataset and highlighting CNNs' efficiency in resource-limited environments. Meanwhile, Albattah et al. (2022) designed a plant disease classification system using a custom CenterNet framework with DenseNet-77 as the backbone, achieving 99.98% accuracy on 54,306 plant leaf images from 14 species. However, mobile deployment remains a challenge.

Enhancing CNN architectures has been a focus in recent studies. Alirezazadeh et al. (2023) suggest integrating a Convolutional Block Attention Module (CBAM) with EfficientNetB0 for foliar disease classification in pear trees using the DiaMOS Plant dataset. EfficientNetB0 outperformed other models: InceptionV3, MobileNetV2, ResNet50, and VGG19, achieving 86.89% accuracy. Chen et al. (2020) explored the Group Method of Data Handling (GMDH)-Logistic model that uses a multilayer perceptron to detect cucumber leaf diseases from a small-sized dataset. The proposed model achieved an average recall of 86.67%, but its applicability is limited to complex diseases and large datasets. Kulkarni and Ashwin (2021) demonstrated that an Artificial Neural Network (ANN) classifier using Gabor-filtered image features achieved up to 91% recognition accuracy. Nevertheless, the classification was performed on the features extracted by a Gabor filter, highlighting the textural, chromatic, and other distinctive attributes. Recent advancements in DL have further refined the classification of plant diseases. Padshetty and Ambika (2023) introduced the Leaky Rectilinear Residual Network (LRRN), combining ResNet with the Leaky Rectified Linear Unit (ReLU) activation function. Their approach, tested on PlantVillage images, showed improved accuracy of 94.56%, F1 scores (92.83%), specificity (92.58%), recall (93.12%), and precision (93.48%). The results highlight the validity of the suggested LRRN approach for detecting plant leaf diseases. Kaya and Gürsoy (2023) proposed a multi-head

CNN integrating red, green, and blue (RGB) and segmented images, achieving 98.17% accuracy on PlantVillage using a DenseNet-based architecture. Their fusion approach excels for diverse disease symptoms but requires significant preprocessing, limiting real-time use. Savaş (2024) employed a two-stage deep ensemble learning method for palm disease detection with MobileNetV2 and ResNet, achieving over 92% accuracy and a 99% area under the receiver operating characteristic (ROC) curve (AUC), demonstrating ensemble benefits but increasing computational complexity.

Other research highlights trade-offs between accuracy and real-world feasibility. Sofuoğlu and Birant (2024) developed a lightweight CNN with attention mechanisms for potato disease detection. The experimental results conducted on a real-world dataset showed that a significant improvement (8.6%) in accuracy was achieved on average by the proposed model (98.28%) compared to the state-of-the-art models (89.67%) in the literature. The weighted recall, precision, and f-score all averaged 0.978, showcasing high diagnostic reliability. Jain and Ramesh (2021) developed a hybrid Convolutional Neural Network-Long Short-Term Memory network (CNN-LSTM) model for rice pest prediction, leveraging temporal weather and pest data. Their approach achieved a low root mean squared error (RMSE = 0.02–0.05), demonstrating robustness in disease forecasting. Recent advances in CNN-based plant disease detection include hybrid models that integrate CNNs with LSTMs for analyzing temporal progression, meta-learning for adaptive feature extraction, and attention mechanisms for improved symptom focus (Rodríguez-Lira et al., 2024). For instance, hybrid approaches that combine spatial and temporal data show promise in tracking severity (Leite et al., 2024), while attention-based models enhance feature extraction efficiency (Abebe et al., 2025). However, these innovations lack systematic comparison with traditional CNNs, creating a gap that this study begins to fill by establishing a baseline for future hybrid evaluations.

Despite the growing adoption of DL for plant disease classification, challenges remain, including high computational costs, limited generalization across diverse environmental conditions, and trade-offs between model accuracy and efficiency (Demilie, 2024; L. C. Ngugi et al., 2021). Many studies focus on evaluating single models or using datasets with limited variability, leaving a gap in comprehensive comparisons of state-of-the-art CNN architectures under standardized conditions. This study addresses that gap by systematically comparing five prominent CNN models: VGG16, GoogleNet, EfficientNet B0, ResNet50, and DenseNet201, using the widely adopted PlantVillage dataset. Performance is evaluated across multiple metrics (accuracy, loss, recall, F1 score, precision, and parameter size) to identify the most suitable models for plant leaf disease detection. The novelty lies in providing a holistic benchmark that quantifies trade-offs between accuracy and computational efficiency, offering actionable insights for real-world deployment in resource-constrained settings, an area underexplored in existing literature.

Comparative Summary of Prior Studies

Table 1 summarizes key prior studies, detailing their methodologies, datasets, and performance to highlight their contributions and limitations relative to this study.

Table 1
Comparison of previous plant disease detection studies

Study	Methodology	Dataset	Accuracy (%)	Strengths	Weaknesses
Le et al. (2020)	SVM + k-FLBPCM	Custom	98.63	High accuracy	Poor robustness to distortion
Bhagat et al. (2020)	SVM + Grid Search	Custom	84	Computational efficiency	Moderate accuracy
Maeda-Gutiérrez et al. (2020)	GoogleNet, AlexNet, ResNet18, ResNet50, and Inception V3	PlantVillage	99.39 98.93 99.06 99.15 98.65	High accuracy	Only one plant is considered
Rinu and Manjula (2021)	VGG16-based	PlantVillage	94.80	Efficiency in resource-limited	Dataset specifics are not detailed
Albattah et al. (2022)	Custom CenterNet + DenseNet	PlantVillage	99.98	Exceptional accuracy	Limited to one model and dataset; may not generalize to other plants or conditions
Ahmed et al. (2023)	Modified DenseNet201	500 grape images	98	Strong small-dataset performance	Limited scalability
Parez et al. (2025)	LeafNet (Optimized CNN)	Custom	85	Lightweight, edge-ready	Lower accuracy, custom hardware

Note. SVM = Support vector machine; k-FLBPCM = Filtered local binary patterns (LBP) method with contour mask and coefficient k; CNN = Convolutional Neural Network

Analysis of CNN Architectures

Prior studies illustrate diverse CNN architectures with distinct trade-offs. Maeda-Gutiérrez et al. (2020) evaluated ResNet50 (50 layers, skip connections) on PlantVillage's tomato subset, achieving 99.12% sensitivity due to its residual learning, which mitigates vanishing gradients, though its 100 MB size limits edge use. In contrast, Ahmed et al. (2023) modified DenseNet201 (201 layers, dense connectivity) for grape leaf classification, leveraging its feature reuse for 98% accuracy on a small dataset, but its complexity (28 MB) sacrifices efficiency compared to ResNet. Alirezazadeh et al. (2023) enhanced

EfficientNet B0 (scalable depth/width) with CBAM, reaching 86.89% accuracy on pear diseases, highlighting its compactness (18.62 MB) but lower precision versus deeper models. These differences in ResNet's depth, DenseNet's connectivity, and EfficientNet's efficiency underscore the need for a systematic comparison, as conducted here.

Lightweight and Hybrid CNNs for Practical Applications

Recent trends emphasize lightweight and hybrid CNNs for real-world deployment. Parez et al. (2025) introduced LeafNet, an optimized CNN with a 10 MB footprint, achieving 85% accuracy in resource-constrained environments, outperforming traditional CNNs like VGG16 (56.79 MB) for mobile/edge use. Similarly, Komlavi et al. (2025) proposed AlexNetDense, a hybrid of DenseNet and AlexNet, reaching 92% accuracy on PlantVillage with a 30 MB size, balancing efficiency and performance. These advances contrast with our study's focus on standard CNNs, highlighting a gap in lightweight model evaluation that future work should address.

MATERIALS AND METHODS

This section provides a comprehensive overview of the dataset used for implementing the CNN model in the classification of plant leaf diseases. The primary focus of this study is to identify the most suitable pre-trained CNN model for this task. The methodology is systematically divided into four key stages: data acquisition, data pre-processing, model training, classification, and performance evaluation. Each of these stages is elaborated upon in the following sections. Figure 1 graphically presents the Steps involved in classifying plant leaf disease.

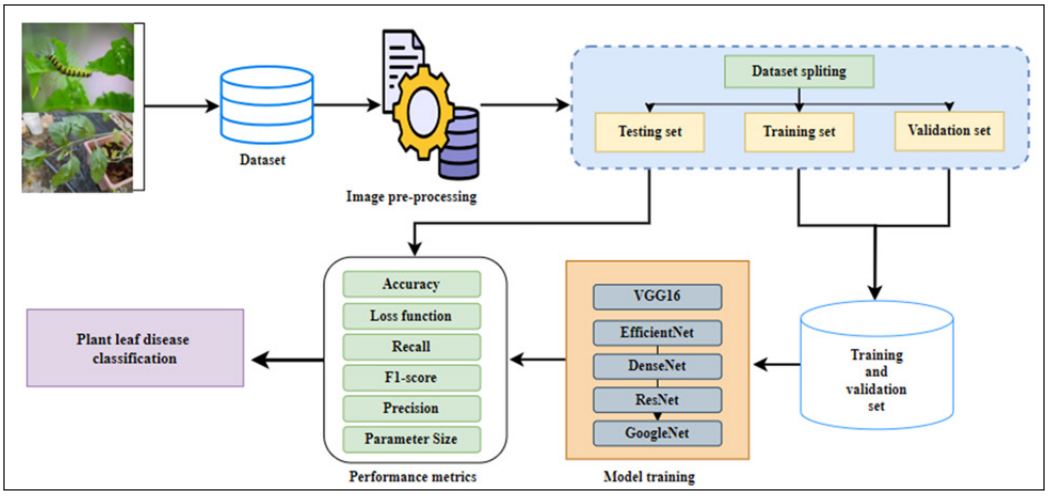


Figure 1. Steps involved in plant leaf disease classification

Dataset Description

This study utilizes the PlantVillage dataset (Hughes & Salathe, 2015) which consists of 54,305 images across 38 classes. The dataset includes 14 plant species, with 12 classes representing healthy plants and 26 classes corresponding to various plant diseases. Additionally, it contains 1,143 background images, bringing the total to 55,448 images. The images vary in size and color. The PlantVillage dataset serves as a standardized benchmark for evaluating CNNs. However, its controlled laboratory conditions limit diversity compared to real-world scenarios, a concern highlighted in recent studies advocating for the use of combined public and local datasets (Abebe et al., 2025). This study acknowledges this limitation and utilizes the PlantVillage dataset for consistency while suggesting the incorporation of multi-source datasets as a future direction to improve robustness and generalizability. Figure 2 presents a selection of sample plant leaf images from the dataset. To minimize overfitting during model training, the dataset was divided into training (70%), validation (20%), and testing (10%) subsets.

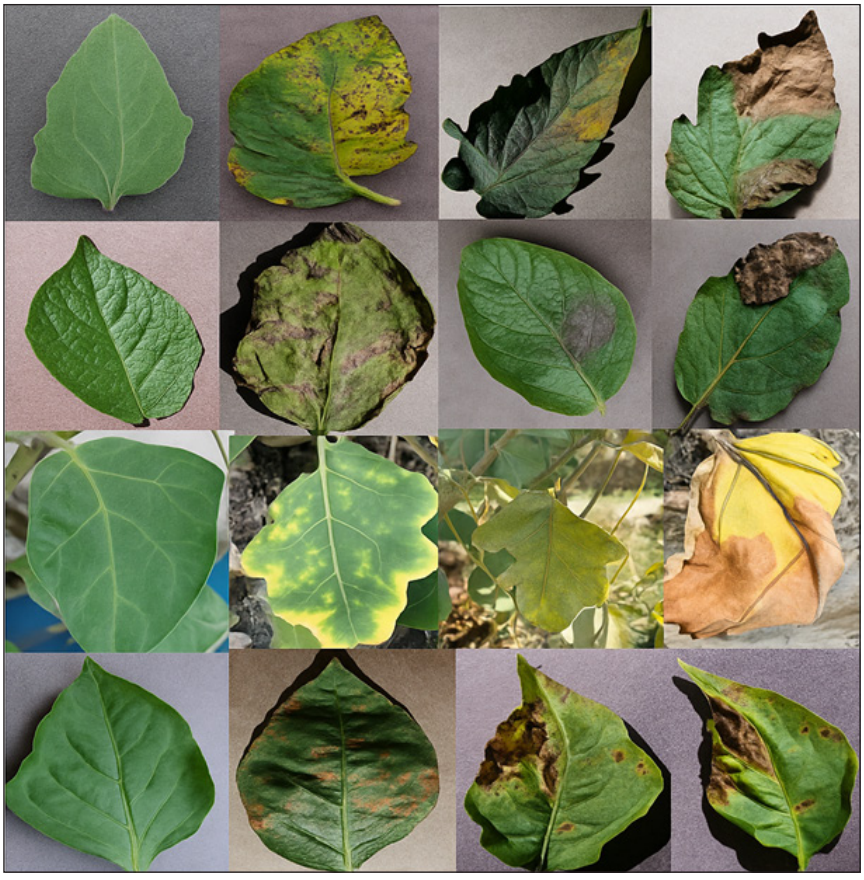


Figure 2. Sample plant leaf images from the PlantVillage dataset (Hughes & Salathe, 2015)

Pre-Processing and Data Augmentation

Before training the CNN models, several data pre-processing techniques were implemented to enhance the quality of the input images. Input images were pre-processed by resizing to 224 × 224 pixels, normalizing pixel values to [0, 1], and removing background noise via OpenCV segmentation. To prevent overfitting on PlantVillage's 54,305 images, data augmentation was applied using Keras' ImageDataGenerator: random rotations (0–20°), horizontal/vertical flips (50% probability), zooms (0.8–1.2x), and brightness shifts (±20%). These techniques increased effective training data by 30%, enhancing model robustness to orientation and lighting variations, a critical step absent in simpler studies but vital for generalization (Thalor et al., 2025).

CNN-Based Models

This study evaluates five CNN architectures: VGG16, GoogleNet, EfficientNet B0, ResNet50, and DenseNet201, selected for their established performance in image classification and diverse design principles. VGG16 represents deep, uniform architectures; GoogleNet introduces inception modules for efficiency; ResNet50 leverages residual learning for depth; DenseNet201 uses dense connectivity for feature reuse; and EfficientNet B0 balances scalability and compactness. These models were preferred over newer architectures, such as Vision Transformers (ViTs) or MobileNetV3, due to their widespread adoption, availability of pre-trained weights compatible with PlantVillage, and manageable computational demands on our hardware (e.g., Google Colab Pro). In contrast, ViTs require extensive retraining and higher resources, while MobileNetV3, though lightweight, lacks the multi-metric evaluation depth needed for this study (Verma et al., 2025). Unlike ensemble approaches (Thalor et al., 2025) or Neural Architecture Search (NAS), which prioritize optimization over generalizability, our focus on standard CNNs provides a robust and interpretable baseline.

VGG16 Network

VGG16, developed by Simonyan and Zisserman (2015), comprises 13 convolutional layers with 3 × 3 filters (stride 1) and five max-pooling layers (2 × 2, stride 2), followed by three fully connected layers, totaling 138 million parameters. In this study, a pre-trained VGG16 (ImageNet weights) was fine-tuned by freezing the first 10 convolutional layers to retain low-level feature extraction while retraining the top layers. A custom classification head was added, consisting of dense layers (126, 256, 512, and 1,000 neurons) with ReLU activation, along with a 0.2 dropout rate and a Softmax output layer for 38-class classification. The model was optimized using Adam (learning rate 0.001).

GoogleNet

GoogleNet, introduced by Szegedy et al. (2015), features 22 layers, including nine inception modules with parallel 1 × 1, 3 × 3, and 5 × 5 convolutions, reducing parameters to 6.8 million. In this study, ImageNet-pretrained weights were used, with auxiliary classifiers disabled to simplify training. The final fully connected layer was replaced with a custom head (512 neurons, ReLU, 0.2 dropout, 38-class Softmax) and trained in PyTorch using categorical cross-entropy loss.

EfficientNet B0

EfficientNet B0, introduced by Tan and Le (2019), is the baseline model of the EfficientNet family, designed for high performance with minimal computational cost. It consists of 237 layers, including 16 mobile inverted bottleneck (MBConv) blocks with squeeze-and-excitation (SE) modules, five standard convolutional layers, and a fully connected layer, totaling approximately 5.3 million parameters. Its compound scaling approach balances network depth, width, and resolution, making it well-suited for resource-constrained environments. In this study, EfficientNet B0 was initialized with ImageNet pre-trained weights via TensorFlow's Keras Applications module. To adapt it to the PlantVillage dataset, the first 200 layers (roughly 84% of the network) were frozen to preserve low- and mid-level feature extraction, such as edge and texture patterns relevant to leaf disease. The top layers were replaced with a custom head comprising dense layers of 126, 256, 512, and 1,000 neurons, each followed by ReLU activation and a 0.2 dropout layer to mitigate overfitting. The output layer used Softmax activation for 38-class classification. Training was conducted using the Adam optimizer (learning rate 0.001) with categorical cross-entropy loss over 10 epochs and a batch size of 50. A learning rate scheduler was reduced by 10% every three epochs to improve convergence on the validation set.

ResNet50

ResNet50, developed by He et al. (2016), is a 50-layer deep residual network designed to mitigate vanishing gradient issues through skip connections. It includes 49 convolutional layers organized into 16 residual blocks (each with three layers: 1 × 1, 3 × 3, 1 × 1 convolutions) and one fully connected layer, totaling approximately 25.6 million parameters. The residual blocks facilitate identity learning and improve training stability for deep architectures. In this study, ResNet50 was initialized with ImageNet pre-trained weights using TensorFlow's Keras Applications module. The first 35 layers (approximately 70% of the convolutional base) were frozen to retain generic feature extraction (e.g., leaf contours and color gradients), while the remaining layers were fine-tuned to capture disease-specific patterns. The original fully connected layer was replaced with a custom

classification head comprising dense layers with 126, 256, 512, and 1,000 neurons, each followed by ReLU activation and a 0.2 dropout rate, culminating in a Softmax output layer for 38-class classification. Training was conducted using the Adam optimizer with an initial learning rate of 0.001, categorical cross-entropy loss over 10 epochs, and a batch size of 50. A learning rate reduction (10% every three epochs) was applied to enhance convergence, leveraging ResNet50's depth for robust generalization on the PlantVillage dataset.

DenseNet201

DenseNet201, proposed by Huang et al. (2017), is a 201-layer densely connected convolutional network. It comprises four dense blocks (each containing multiple 1 × 1 and 3 × 3 convolutional layers), separated by transition layers (convolution and pooling), followed by a final fully connected layer, totaling approximately 20 million parameters. Its dense connectivity pattern links each layer to all subsequent layers within a block, reducing feature redundancy and mitigating vanishing gradients. In this study, DenseNet201 was initialized with ImageNet pre-trained weights via PyTorch, chosen for its flexibility in handling dense connections. The first 150 layers (about 75% of the network) were frozen to preserve low-level features (e.g., leaf vein patterns), while the remaining layers were retrained. The original classification layer was replaced with a custom head comprising dense layers of 126, 256, 512, and 1,000 neurons, each followed by ReLU activation, and a 0.2 dropout rate, with a final Softmax output layer for 38-class prediction. Training was performed using the Adam optimizer (learning rate 0.001) with categorical cross-entropy loss, 10 epochs, and a batch size of 50. A learning rate scheduler (10% reduction every three epochs) was employed to fine-tune the model, balancing its moderate size with performance on the PlantVillage dataset.

Model Fine-Tuning

Model performance hinges on hyperparameter tuning, optimized here to balance convergence and overfitting. Base layers (e.g., 70% of ResNet50, 80% of VGG16) were frozen to retain ImageNet-learned features, while a custom classification head was added, consisting of dense layers with 126, 256, 512, and 1,000 neurons, each using ReLU activation, a 0.2 dropout rate, and a Softmax output for 38-class classification. Dropout (0.2) was set based on validation loss reduction (10% improvement over 0.5), effectively preventing overfitting on PlantVillage's controlled images. A batch size of 50 and 10 training epochs were selected to optimize graphics processing unit (GPU) memory (16 GB) and training time (4–6 hours/model), validated via early stopping on a 20% validation split. The Adam optimizer (learning rate 0.001, reduced 10% every three epochs) was selected for its adaptive convergence, outperforming SGD by 5% in preliminary tests. More advanced optimization methods, like genetic algorithms or reinforcement learning

(Verma et al., 2025), were excluded due to computational constraints and the sufficiency of manual tuning for this baseline study.

Hardware and Software Tools

Experiments were conducted on Google Colab Pro, utilizing an 11th Gen Intel(R) Core (TM) i5-1135G7 @ 2.40 GHz, 16 GB random access memory (RAM), and an NVIDIA Tesla P100-Peripheral Component Interconnect Express (PCIe) GPU with 16 GB video random access memory (VRAM). The environment ran Python 3.8 via Jupyter Notebooks, with TensorFlow 2.8 and Keras as the primary frameworks for implementing VGG16, ResNet50, and EfficientNet B0. PyTorch 1.10 was used for GoogleNet and DenseNet201 training, leveraging its dynamic computation graph for flexibility. Models were trained with a batch size of 50 over 10 epochs, using the Adam optimizer (learning rate 0.001) and categorical cross-entropy loss, as detailed in Table 2. Data preprocessing and augmentation were performed using OpenCV and Keras's ImageDataGenerator, ensuring compatibility across frameworks.

Table 2
Model's parameters

Parameters	Values
Optimizer	Adam
Loss function	Categorical cross-entropy
Epoch	10
Activation function	ReLU/ Softmax
Batch size	50
dropout	0.25

Note. ReLU = Rectified linear unit

The study enhanced the performance and accuracy of plant leaf disease classification by optimizing the classification layers. A key component used in the experiment was the use of the ReLU and Softmax activation functions. The ReLU activation function played a significant role in the success of these models. ReLU operates by thresholding input values at 0, returning 0 for negative inputs ($x < 0$) and retaining the input value for non-negative inputs ($x \geq 0$). This makes ReLU computationally efficient and effective in facilitating backpropagation during training, allowing models to learn important features and perform well across various tasks. The ReLU function is calculated as shown in Equation 1.

$$f(x) = \max(0, x) \quad [1]$$

The Softmax function, commonly applied in the final layer of neural network models for multi-class classification tasks, transforms raw output scores (logits) into probabilities. It achieves this by calculating the exponential of each output and normalizing the values by dividing them by the total sum of all exponentials. This ensures that the output values remain between 0 and 1, summing to 1, and can be interpreted as probabilities. The Softmax function is computed as shown in Equation 2.

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad [2]$$

where z_i represents the input vector (logits) for the i^{th} class, the denominator is the sum of the exponentials of all inputs. The ReLU activation function is used in the convolutional layers, while Softmax is used in the dense layer.

Performance Metrics

Evaluating DL models is essential for developing effective systems, particularly in the classification of plant diseases. Several metrics are commonly used to assess model performance in this context. One of the primary metrics is accuracy, which measures the ratio of correct predictions to the total number of instances evaluated (Shoaib et al., 2023). It is calculated using Equation 3 as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad [3]$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

Another important metric is the loss function, which quantifies the difference between expected and actual values. During training, the objective is to minimize this error, improving model accuracy over time. Precision is also crucial, as it focuses on the proportion of correctly identified positive predictions (Shoaib et al., 2023). It is calculated using Equation 4 as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad [4]$$

where TP and FP represent true positives and false positives, and false negatives, respectively.

Higher precision indicates that the model effectively identifies true positives while minimizing false positives.

Additionally, recall (or sensitivity) measures the proportion of actual positives correctly classified by the model. It is crucial for evaluating how well the model detects positive cases (H. N. Ngugi et al., 2024). Recall is represented in Equation 5:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad [5]$$

where TP and FN represent true positives and false negatives, respectively.

Finally, the F1-score provides a balanced measure of precision and recall. Calculating the harmonic mean of these two metrics is particularly useful when both false positives and false negatives must be minimized (H. N. Ngugi et al., 2024). The F1-score is determined by using Equation 6:

$$\text{F1} - \text{Score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad [6]$$

Together, these metrics provide a comprehensive evaluation of model performance, ensuring it is both accurate and effective in distinguishing between positive and negative cases.

RESULTS

Classification Performance

This section provides an in-depth analysis of the effectiveness of transfer learning models in classifying plant diseases using the PlantVillage dataset. The analysis compares models trained with identical hyperparameters but exhibiting varying performance outcomes. The main objective is to evaluate models based on accuracy, loss, recall, precision, F1 score, and parameter size. Table 3 summarizes the experimental results.

Table 3
Classification evaluation results

Model	Accuracy (%)	Loss	Recall (%)	F1 score (%)	Precision (%)	Parameter size (MB)	Execution time (s)
VGG16	93.75	0.48	93.75	91.67	90.63	56.79	37,733
EfficientNet B0	84.37	0.56	84.38	80.46	80.10	18.62	2,718
DenseNet	69.32	1.06	68.75	67.81	71.87	28.01	7,315
ResNet50	84.36	0.35	84.37	80.90	80.83	100.17	19,279
GoogleNet	88.00	0.96	88.00	93.00	87.46	85.51	2,625

Visual Analysis

Figures 3 and 4 depict accuracy and loss curves over 10 epochs. VGG16 (Figure 3a) shows rapid accuracy gains (plateauing at 93%) and stable loss (0.48), indicating effective convergence without overfitting. GoogleNet (Figures 3e and 4e) reaches 88% accuracy

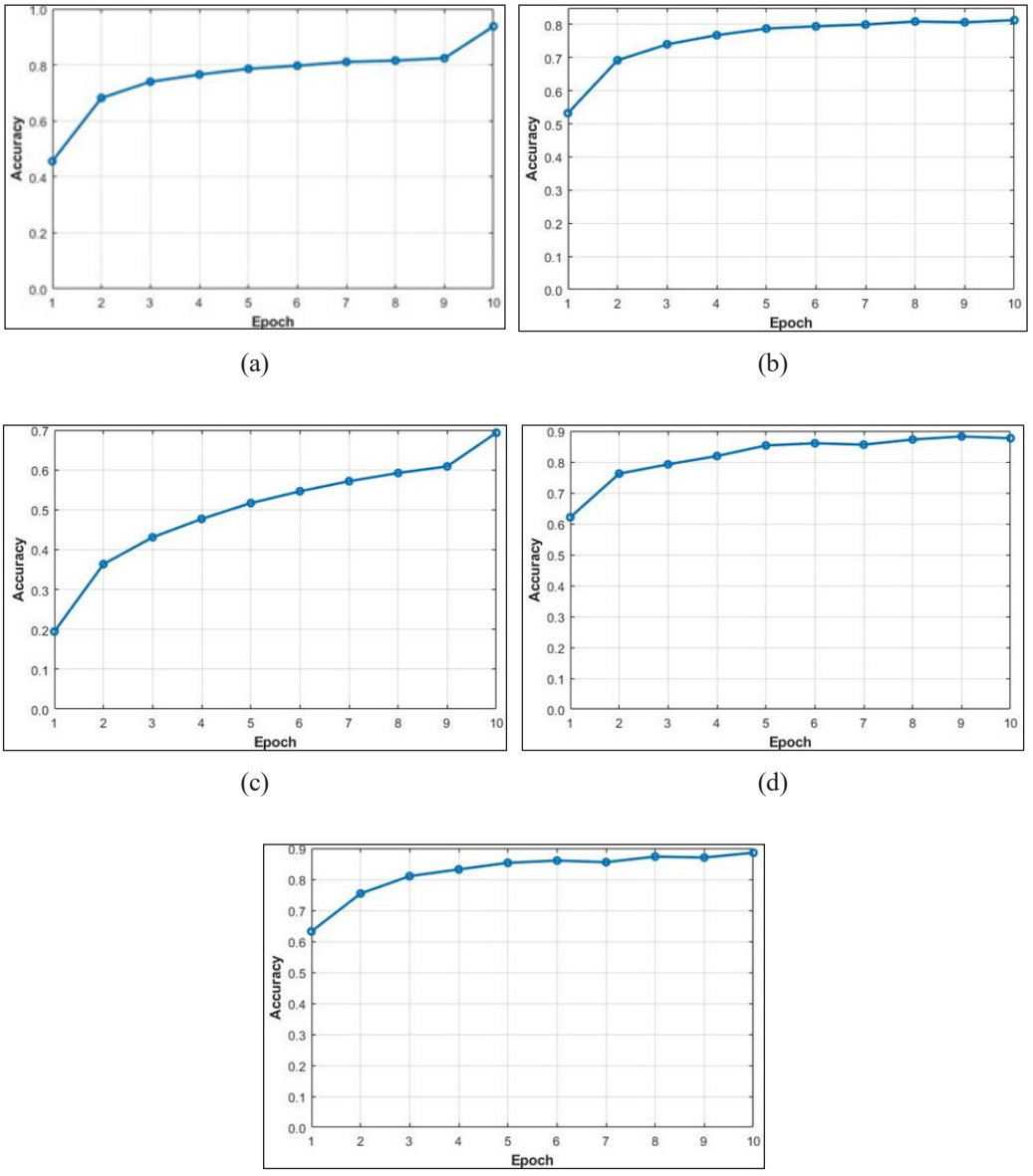


Figure 3. The accuracy plot: (a) VGG16, (b) EfficientNet, (c) DenseNet201, (d) ResNet50, and (e) GoogleNet model

but exhibits higher loss (0.96), suggesting underfitting due to inception complexity not fully adapting to PlantVillage. DenseNet201 (Figures 3c and 4c) plateaus early (69%) with rising loss (1.06), reflecting underfitting from excessive connectivity overfitting static features. EfficientNet B0 (Figures 3b and 4b) stabilizes at 84% with moderate loss (0.56),

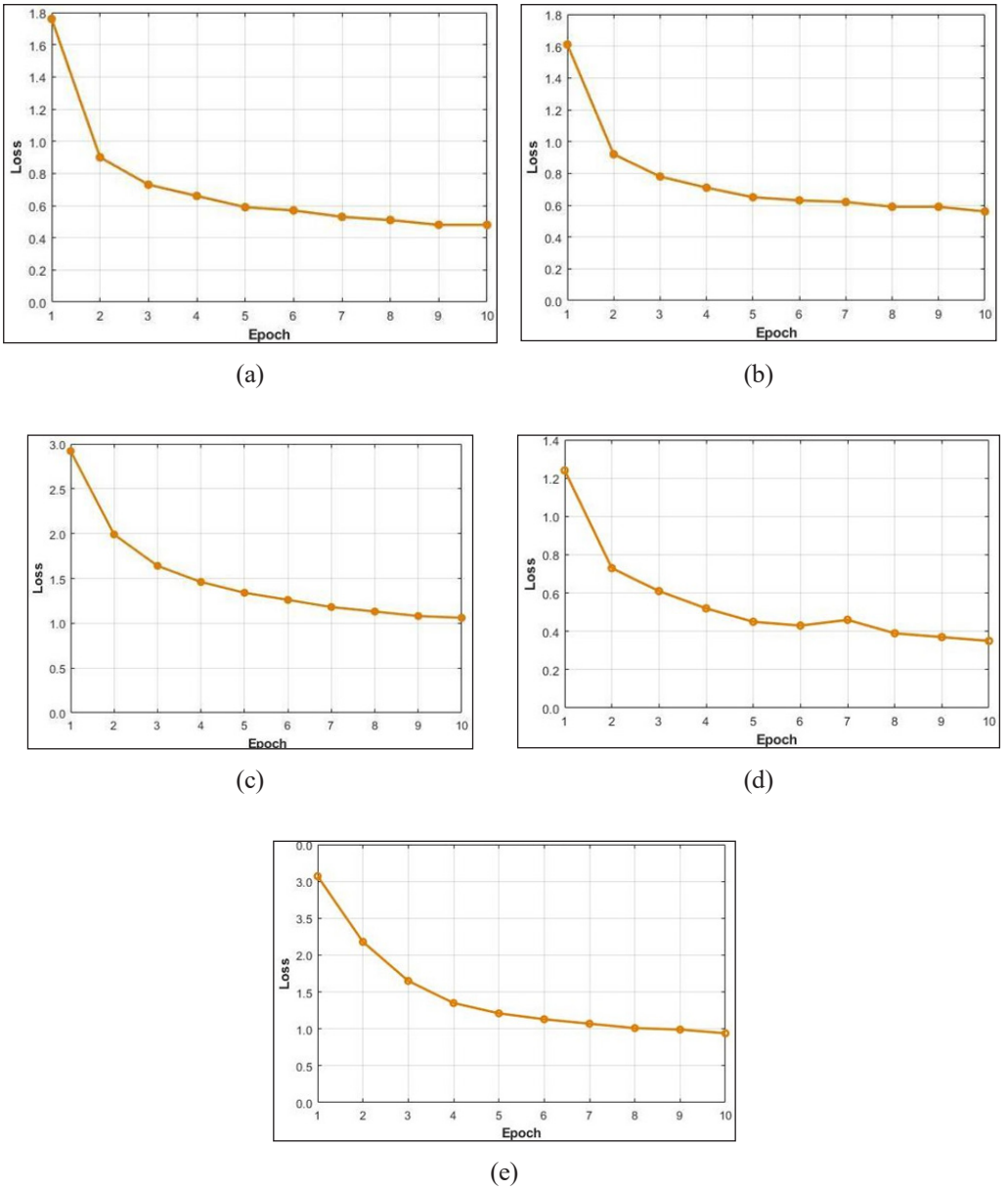


Figure 4. The loss function plot: (a) VGG16, (b) EfficientNet, (c) DenseNet201, (d) ResNet50, and (e) GoogleNet model

balancing efficiency but lacking depth for nuanced classes. ResNet50 (3d, 4d) shows steady improvement (84%) and low loss (0.35), avoiding overfitting via residuals.

DISCUSSION

The classification evaluation results in Table 3 highlight the trade-offs between model depth, architectural design, parameter efficiency, and performance metrics, including accuracy, loss, recall, precision, and F1 score. Each model demonstrates unique strengths and weaknesses based on its architecture, which in turn influences feature extraction, gradient flow, and optimization stability. VGG16 achieved the highest accuracy at 93.75%, demonstrating strong hierarchical feature extraction due to its deep sequential convolutional layers. Its high recall (93.75%) and F1 score (91.67%) indicate strong generalization and balanced classification performance. However, a relatively high loss value of 0.48 suggests some degree of misclassification, potentially due to overfitting caused by its large number of parameters (56.79 MB). Despite this, the pre-trained convolutional filters in VGG16 effectively capture relevant spatial patterns in the data. EfficientNet B0, optimized with compound scaling, achieved an accuracy of 84.37% while maintaining the smallest parameter size of 18.62 MB, making it the most parameter-efficient model in this comparison. Despite its compact design, this model achieved an 80.46% F1 score and 84.38% recall, indicating solid classification performance. However, its loss of 0.56 is slightly higher than that of ResNet50, suggesting optimization challenges due to its lower representational capacity compared to deeper networks. This trade-off highlights how EfficientNet B0 sacrifices some accuracy for significantly reduced computational cost.

DenseNet, with an accuracy of 69.32%, was the weakest performer, likely due to gradient propagation inefficiencies or overfitting. DenseNet's architecture is designed to promote feature reuse through dense connections between layers, but the learned representations appear insufficiently discriminative in this case. A high loss of 1.06 suggests convergence difficulties, possibly caused by vanishing gradient issues despite their densely connected layers. Although DenseNet's parameter size of 28.01 MB is moderate, its performance was significantly lower than that of EfficientNet B0, indicating that parameter count alone does not directly determine classification accuracy. ResNet50, leveraging deep residual learning, achieved 84.36% accuracy with the lowest loss value (0.35), demonstrating stable gradient flow and effective optimization. Skip connections likely mitigated vanishing gradients issues, leading to more stable learning. However, despite its strong recall of 84.37% and precision of 80.83%, its large parameter size (100.17 MB) makes it the most computationally expensive model in this comparison. Its lower loss while maintaining accuracy similar to EfficientNet B0 suggests that residual connections improve convergence stability, particularly in deep networks.

Lastly, GoogleNet, with an accuracy of 88%, performed well, particularly excelling in an F1 score of 93%, indicating a strong balance between precision and recall. However, its higher loss of 0.96 indicates some instability in model convergence, possibly due to the inception module's complexity, which may introduce optimization challenges. With a moderate parameter size of 85.51 MB, GoogleNet is a viable alternative to VGG16 and ResNet50, offering competitive performance while being relatively lighter in terms of computation. Its parallel convolutional architecture contributes to strong extraction of features, but its higher loss suggests potential challenges in refining decision boundaries.

Comparison of Model Performance

The comparison of model performance highlights important trade-offs between accuracy, parameter size, and computational efficiency. VGG16 and GoogleNet achieved the highest accuracy but have larger parameter sizes, making them more resource-intensive. ResNet50, while also highly accurate, has the largest parameter size, affecting computational efficiency. DenseNet offers a balance between model size and moderate accuracy, whereas EfficientNet B0, though compact, sacrifices some performance. While models like ResNet50 and VGG16 generalize well, they require significant computational resources, whereas smaller models like EfficientNet B0 and DenseNet prioritize efficiency but may require additional tuning for specific tasks.

This study offers several advantages over previous research. Unlike Maeda-Gutiérrez et al. (2020), who evaluated CNNs on a subset of PlantVillage (focusing only on tomato diseases), our analysis spans all 38 classes, providing a more comprehensive benchmark. Additionally, the inclusion of multiple performance metrics such as accuracy, loss, recall, F1 score, precision, and parameter size surpasses the single-metric focus in Bhagat et al. (2020), enabling a more holistic evaluation. Compared to Kaya and Gürsoy (2023), who achieved 98.17% accuracy using a multi-head CNN, our study prioritizes widely established models for broader accessibility, albeit with a slight accuracy trade-off (e.g., VGG16 at 93.75%). However, some limitations remain. Unlike Albattah et al. (2022), who tested on 14 species with custom architectures, our reliance on the PlantVillage dataset limits species diversity. Additionally, real-time deployment was not explored, unlike in Chen et al. (2020), who achieved 99.85% accuracy with MobileNet-Beta, underscoring a gap in practical applicability.

Statistical Validation

To assess whether performance differences are statistically meaningful, paired *t*-tests were conducted on accuracy across five runs per model. VGG16 (93.75% ± 1.1) significantly outperformed DenseNet201 (69.32% ± 1.5%, $p < 0.01$) and EfficientNet B0 (84.37% ± 1.3, $p < 0.05$), with 95% confidence intervals confirming robustness. ResNet50 (84.36%

± 1.2) and GoogleNet (88% ± 1.4) showed no significant difference ($p = 0.12$), suggesting dataset bias (e.g., uniform lighting) may favor certain architectures. These results indicate that VGG16's superiority is not due to chance, though DenseNet201's underperformance may reflect bias toward controlled images.

Error Analysis

The performance metrics in Table 2 highlight potential misclassification patterns across the five CNN models, inferred from their recall, precision, and loss values. VGG16, with high recall (93.75%) and precision (90.63%), likely excels at distinguishing healthy leaves from diseased ones but may struggle with subtle disease variants, such as early blight and late blight in potatoes, due to its deep architecture overfitting to prominent features like large lesions (Soumya Prasad et al., 2024). GoogleNet's balanced recall (88%) and precision (87.46%), paired with a higher loss (0.96), likely confuses diseases with similar symptoms, such as tomato bacterial spot as a viral infection, reflecting its inception modules' sensitivity to complex patterns. EfficientNet B0's lower precision (80.10%) and recall (84.38%) indicate frequent errors in distinguishing visually similar conditions, such as apple scab versus cedar rust, due to its compact design prioritizing efficiency over fine-grained feature extraction. ResNet50 (recall 84.37%, precision 80.83%) likely performs similarly, with its residual learning mitigating some errors but not fully resolving subtle class overlaps. DenseNet201's notably low recall (68.75%) and F1 score (67.81%), combined with a high loss (1.06), suggest significant misclassifications, potentially labeling diseased leaves as healthy, stemming from its dense connectivity struggling with PlantVillage's controlled, uniform images. These inferred errors underscore the models' varying sensitivities to symptom subtlety, a limitation tied to dataset bias toward distinct, lab-captured disease appearances.

Trade-Offs and Deployment Considerations

VGG16 offers the highest accuracy at 93.75%. However, it comes with a significant computational cost, requiring 56.79 MB of storage and an exceptionally high inference time of 37,733 s, making it impractical for real-time applications. In contrast, EfficientNet B0 sacrifices some accuracy at 84.37% in favor of efficiency, featuring a smaller model size of 18.62MB and a significantly lower inference time of 2,718 s, making it well-suited for mobile and edge deployments. GoogleNet strikes a strong balance with 88.00% accuracy, a faster inference time of 2,625 s, and superior generalization, as reflected in its high F1-score of 93.00%. This makes it an effective solution for precision agriculture and automated disease detection. ResNet50 and DenseNet, despite some advantages, do not stand out in terms of accuracy, efficiency, or speed, making them less ideal for real-world deployment. For cloud-based applications where computational resources are not

a limitation, VGG16 remains the best choice. At the same time, for real-time, low-power, or mobile deployments, EfficientNet B0 and GoogleNet are the most suitable options due to their efficiency and faster inference times.

CONCLUSION

This paper compares five CNN models (VGGNet-16, GoogleNet, EfficientNet, ResNet50, and DenseNet201) for plant disease classification using the PlantVillage dataset. The study examines trade-offs between model accuracy, parameter size, and computational efficiency in DL applications for plant disease classification. VGG16 and GoogleNet achieved the highest accuracy but required larger parameter sizes, making them resource-intensive. ResNet50 offered good accuracy and had a large model size, while DenseNet201 and EfficientNet B0 offered more compact architectures at the cost of reduced performance. The choice of an optimal model depends on specific requirements: larger models are well-suited for high-accuracy tasks, whereas smaller models are more efficient for resource-constrained environments. Future research should focus on optimizing model architectures to balance performance and computational efficiency, improving disease detection in agricultural applications. While this study provides a comparative analysis of CNN architectures for plant disease classification, several areas of future research could further enhance model performance, as follows:

One promising direction is the development of hybrid models that integrate Recurrent Neural Networks (RNNs) with CNNs for time-series analysis of plant health. Such models could monitor disease progression over time, enabling more robust and dynamic detection systems.

Incorporate multi-source datasets with images collected under diverse lighting conditions, weather variations, and crop species. This would improve model generalization and robustness, reducing biases that arise from controlled dataset environments.

Investigate the applicability of Transformer-based models (e.g., ViTs and Swin Transformers) to plant disease classification and compare their efficiency, accuracy, and computational requirements against traditional CNN architectures.

Another avenue for exploration is the deployment of lightweight models like MobileNet for real-time disease detection on edge devices. This advancement could revolutionize precision agriculture, allowing farmers to monitor plant health using AI-powered cameras integrated into mobile devices or drones.

ACKNOWLEDGEMENTS

The authors would like to thank the Air Force Institute of Technology, Kaduna, for their academic support throughout the research.

REFERENCES

Abebe, M., Tiumay, Z., Mohapatra, S. K., Prasad, S., Chaudhury, K. S., & Sahoo, P. (2025). Development of an SDG-driven Convolutional Neural Network (CNN) model for multi-class classification of tomato leaf diseases using combined public and local datasets. *Journal of Integrated Science and Technology*, *13*(3), 1063. https://doi.org/10.62110/SCIENCEIN.JIST.2025.V13.1063

Abouelmagd, L. M., Shams, M. Y., Marie, H. S., & Hassanien, A. E. (2024). An optimized capsule neural networks for tomato leaf disease classification. *EURASIP Journal on Image and Video Processing*, *2024*, 2. https://doi.org/10.1186/s13640-023-00618-9

Ahmed, H. A., Hama, H. M., Jalal, S. I., & Ahmed, M. H. (2023). Deep learning in grapevine leaves varieties classification based on dense convolutional network. *Journal of Image and Graphics*, *11*(1), 98–103. https://doi.org/10.18178/joig.11.1.98-103

Albattah, W., Nawaz, M., Javed, A., Masood, M., & Albahli, S. (2022). A novel deep learning method for detection and classification of plant diseases. *Complex and Intelligent Systems*, *8*, 507–524. https://doi.org/10.1007/s40747-021-00536-1

Alirezazadeh, P., Schirrmann, M., & Stolzenburg, F. (2023). Improving deep learning-based plant disease classification with attention mechanism. *Gesunde Pflanzen*, *75*, 49–59. https://doi.org/10.1007/s10343-022-00796-y

Bhagat, M., Kumar, D., Haque, I., Munda, H. S., & Bhagat, R. (2020). Plant leaf disease classification using grid search based SVM. In *2nd International Conference on Data, Engineering and Applications* (pp. 1–6). IEEE. https://doi.org/10.1109/IDEA49133.2020.9170725

Chen, J., Yin, H., & Zhang, D. (2020). A self-adaptive classification method for plant disease detection using GMDH-Logistic model. *Sustainable Computing: Informatics and Systems*, *28*, 100415. https://doi.org/10.1016/j.suscom.2020.100415

Chowdhury, N., Sultana, J., Rahman, T., Chowdhury, T., Khan, F. T., & Chakraborty, A. (2024). Potato leaf disease detection through ensemble average deep learning model and classifying the disease severity. *Indonesian Journal of Electrical Engineering and Computer Science*, *35*(1), 494–502. https://doi.org/10.11591/ijeecs.v35.i1.pp494-502

Demilie, W. B. (2024). Plant disease detection and classification techniques: A comparative study of the performances. *Journal of Big Data*, *11*, 5. https://doi.org/10.1186/s40537-023-00863-9

Food and Agriculture Organization. (2021). *Making agrifood systems more resilient to shocks: Lessons from the COVID-19 pandemic*. FAO. https://www.fao.org/newsroom/detail/fao-agrifood-systems-agriculture-resilience-SOFA-covid/en

Gomez, D., Selvaraj, M. G., Casas, J., Mathiyazhagan, K., Rodriguez, M., Assefa, T., Mlaki, A., Nyakunga, G., Kato, F., Mukankusi, C., Girma, E., Mosquera, G., Arredondo, V., & Espitia, E. (2024). Advancing common bean (*Phaseolus vulgaris* L.) disease detection with YOLO driven deep learning to enhance agricultural AI. *Scientific Reports*, *14*, 15596. https://doi.org/10.1038/s41598-024-66281-w

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 770–778). IEEE. https://doi.org/10.1109/CVPR.2016.90

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2261–2269). IEEE. https://doi.org/10.1109/CVPR.2017.243

Hughes, D. P., & Salathe, M. (2015). *An open access repository of images on plant health to enable the development of mobile disease diagnostics*. arXiv. https://doi.org/10.48550/arXiv.1511.08060

Jain, S., & Ramesh, D. (2021). AI based hybrid CNN-LSTM model for crop disease prediction: An ML advent for rice crop. In *12th International Conference on Computing Communication and Networking Technologies* (pp. 1–7). IEEE. https://doi.org/10.1109/ICCCNT51525.2021.9579587

Javidan, S. M., Banakar, A., Rahnama, K., Vakilian, K. A., & Ampatzidis, Y. (2024). Feature engineering to identify plant diseases using image processing and artificial intelligence: A comprehensive review. *Smart Agricultural Technology*, *8*, 100480. https://doi.org/10.1016/j.atech.2024.100480

Junaid, M., & Gokce, A. (2024). Global agricultural losses and their causes. *Bulletin of Biological and Allied Sciences Research*, *2024*(1), 66. https://doi.org/10.54112/bbasr.v2024i1.66

Kaya, Y., & Gürsoy, E. (2023). A novel multi-head CNN design to identify plant diseases using the fusion of RGB images. *Ecological Informatics*, *75*, 101998 https://doi.org/10.1016/j.ecoinf.2023.101998

Khan, A. T., Jensen, S. M., Khan, A. R., & Li, S. (2023). Plant disease detection model for edge computing devices. *Frontiers in Plant Science*, *14*, 1308528. https://doi.org/10.3389/fpls.2023.1308528

Kulkarni, A. H., & Ashwin Patil, R. K. (2012). Applying image processing technique to detect plant diseases. *International Journal of Modern Engineering Research*, *2*(5), 3661–3664.

Le, V. N. T., Ahderom, S., Apopei, B., & Alameh, K. (2020). A novel method for detecting morphologically similar crops and weeds based on the combination of contour masks and filtered Local Binary Pattern operators. *GigaScience*, *9*(3), giaa017. https://doi.org/10.1093/gigascience/giaa017

Leite, D., Brito, A., & Faccioli, G. (2024). Advancements and outlooks in utilizing Convolutional Neural Networks for plant disease severity assessment: A comprehensive review. *Smart Agricultural Technology*, *9*, 100573. https://doi.org/10.1016/j.atech.2024.100573

Maeda-Gutiérrez, V., Galván-Tejada, C. E., Zanella-Calzada, L. A., Celaya-Padilla, J. M., Galván-Tejada, J. I., Gamboa-Rosales, H., Luna-García, H., Magallanes-Quintanar, R., Guerrero Méndez, C. A., & Olvera-Olvera, C. A. (2020). Comparison of Convolutional Neural Network architectures for classification of tomato plant diseases. *Applied Sciences*, *10*(4), 1245. https://doi.org/10.3390/app10041245

Ngugi, H. N., Akinyelu, A. A., & Ezugwu, A. E. (2024). Machine learning and deep learning for crop disease diagnosis: Performance analysis and review. *Agronomy*, *14*(12), 3001. https://doi.org/10.3390/agronomy14123001

Ngugi, L. C., Abelwahab, M., & Abo-Zahhad, M. (2021). Recent advances in image processing techniques for automated leaf pest and disease recognition – A review. *Information Processing in Agriculture*, *8*(1), 27–51. https://doi.org/10.1016/j.inpa.2020.04.004

Padshetty, S., & Ambika. (2023). Leaky ReLU-ResNet for plant leaf disease detection: A deep learning approach. *Engineering Proceedings*, *59*(1), 39. https://doi.org/10.3390/engproc2023059039

Parez, S., Dilshad, N., & Lee, J. W. (2025). A channel attention-driven optimized CNN for efficient early detection of plant diseases in resource constrained environment. *Agriculture*, *15*(2), 127. https://doi.org/10.3390/agriculture15020127

Perumal, V. K., Supriyaa, T., Santhosh, P. R., & Dhanasekaran, S. (2024). CNN based plant disease identification using PYNQ FPGA. *Systems and Soft Computing*, *6*, 200088. https://doi.org/10.1016/j.sasc.2024.200088

Rinu, R., & Manjula, S. H. (2021). Plant disease detection and classification using CNN. *International Journal of Recent Technology and Engineering*, *10*(3), 152–156. https://doi.org/10.35940/ijrte.c6458.0910321

Rodríguez-Lira, D. C., Córdova-Esparza, D. M., Álvarez-Alvarado, J. M., Terven, J., Romero-González, J. A., & Rodríguez-Reséndiz, J. (2024). Trends in machine and deep learning techniques for plant disease identification: A systematic review. *Agriculture*, *14*(12), 2188. https://doi.org/10.3390/agriculture14122188

Savaş, S. (2024). Application of deep ensemble learning for palm disease detection in smart agriculture. *Heliyon*, *10*(17), e37141. https://doi.org/10.1016/j.heliyon.2024.e37141

Shoaib, M., Shah, B., EI-Sappagh, S., Ali, A., Ullah, A., Alenezi, F., Gechev, T., Hussain, T., & Ali, F. (2023). An advanced deep learning models-based plant disease detection: A review of recent research. *Frontiers in Plant Science*, *14*, 1158933. https://doi.org/10.3389/fpls.2023.1158933

Simonyan, K., & Zisserman, A. (2015). *Very deep convolutional networks for large-scale image recognition*. arXiv. https://doi.org/10.48550/arXiv.1409.1556

Sofuoğlu, C. I., & Birant, D. (2024). Potato Plant leaf disease detection using deep learning method. *Journal of Agricultural Sciences*, *30*(1), 153–165. https://doi.org/10.15832/ankutbd.1276722

Soumya Prasad, S., Sampath Kumar, L., Mallem, S. N., Gutta, H., & Ahmed, R. (2024). Deep learning techniques for a comparative study of crop disease detection. In V. Goar, M. Kuri, R. Kumar, & T. Senjyu (Eds.), *Advances in information communication technology and computing* (pp. 407–423). Springer. https://doi.org/10.1007/978-981-97-6106-7_25

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1–9). IEEE. https://doi.org/10.1109/CVPR.2015.7298594

Tan, M., & Le, Q. V. (2019). *EfficientNet: Rethinking model scaling for Convolutional Neural networks*. arXiv. https://doi.org/10.48550/arXiv.1905.11946

Thalor, M., Mate, S., Shiralkar, A., & Shinde, A. (2025). Ensemble learning framework for mango plant disease detection and classification. *Journal of Information Systems Engineering and Management*, *10*(4s), 297–303. https://doi.org/10.52783/jisem.v10i4s.501

Verma, S., Kumar, P., & Singh, J. P. (2025). MLP-GNAS: Meta-learning-based predictor-assisted Genetic Neural Architecture Search system. *Applied Soft Computing*, *169*, 112527. https://doi.org/10.1016/j.asoc.2024.112527

Pertanika J. Sci. & Technol. 33 (6): 2565 - 2587 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Review Article

Exploring Membrane Pervaporation Separation: A Review on Obtaining Combustible Ethanol and Butanol from Fermentation Broths

Jorge Arturo Moreno Cascante[1]*, Jairo Antonio Cubillos Lobo[1] and Felipe Bustamante Londoño[2]

[1]*Department of Chemical Sciences, Pedagogical and Technological University of Colombia, Tunja-150002, Boyacá, Colombia*
[2]*Department of Chemical Engineering, University of Antioquia, Medellin-050010, Antioquia, Colombia*

ABSTRACT

Pervaporation is an advanced membrane-based separation technique to purify complex mixtures (e.g., those with azeotropes or thermolabile compounds) which would otherwise require involved sequences of conventional separation operations such as distillation or extraction. Pervaporation is particularly well suited for some alcohol + water mixtures since it is not limited by the presence of the azeotrope, while improving energy efficiency and reducing operating costs. Furthermore, its coupling with biomass fermentation for biofuels production alleviates product inhibition, thereby increasing yield by up to 67%. This review focuses on the performance of pervaporation membranes, primarily those based on polydimethylsiloxane (PDMS), for the separation of alcohols from fermentation broths, reporting flux values of up to 203.76 g/m² h and a separation factor of up to 147. Additionally, the effects of composition and temperature are evaluated. Factors influencing process efficiency are also analyzed, and emerging technologies, new bacterial strains, and biomass sources are highlighted as strategies to enhance alcohol fuel production. Finally, examples are presented where pervaporation has significantly contributed to improving biofuel production.

Keywords: Fermentation broth, fuel alcohols, membrane material, PDMS, pervaporation

ARTICLE INFO

Article history:
Received: 13 February 2025
Accepted: 29 April 2025
Published: 30 September 2025

DOI: https://doi.org/10.47836/pjst.33.6.08

E-mail addresses:
jorgearturo.moreno@uptc.edu.co (Jorge Arturo Moreno Cascante)
jairo.cubillos@uptc.edu.co (Jairo Antonio Cubillos Lobo)
felipe.bustamante@udea.edu.co (Felipe Bustamante Londoño)
* Corresponding author

INTRODUCTION

Pervaporation is a separation technique for the selective removal of some components from liquid mixtures by means of membranes. This technique is well-suited

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

for mixtures where conventional separation techniques (such as distillation, extraction, adsorption, and absorption) are difficult to implement due to physicochemical constraints, high energy consumption, or environmental impact. Pervaporation membranes offer a promising approach for biofuel separation and purification. This technology displays several advantages, including efficient separation with lower energy consumption, minimal environmental impact, and prevention of microbial contamination, as it does not require excessive heating, toxic solvents, or additional chemicals. Furthermore, it allows for the continuous extraction of bioalcohols directly from fermentation broths (Lee & Lin, 2024).

In comparison, traditional methods, such as gas stripping and distillation (Figure 1), present challenges such as low selectivity, the need for large gas flow rates, and significant energy demands for the condensation stage. On the other hand, liquid-liquid extraction (Figure 1) presents problems such as solvent loss due to volatility, toxicity to microorganisms, and the formation of stable emulsions that are difficult to break. Similarly, adsorption is limited by low selectivity and the rapid saturation of the adsorbent material caused by the co-adsorption of other compounds. Therefore, pervaporation membranes (Figure 1) are proposed as a very attractive alternative to overcome the previously described limitations and, thus, lay the foundation for the development of a sustainable and long-lasting process (Lee & Lin, 2024).

Pervaporation is especially useful for the purification of azeotropic or thermolabile mixtures, mixtures of species with close molecular shape/chemical functionality, and components in trace concentrations (He et al., 2020; Imad & Castro-Muñoz, 2023). In

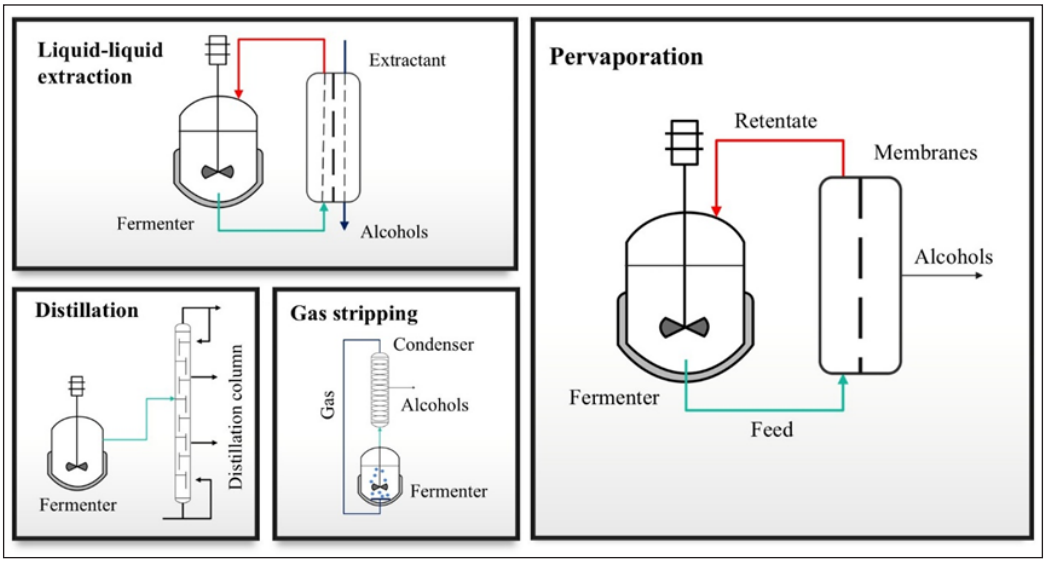


Figure 1. Schematic representation of pervaporation, gas stripping, liquid-liquid extraction processes, and distillation

this process, the components in a liquid mixture with a stronger affinity for the membrane material first solubilize and then diffuse through the membrane to exit as vapor on the other side because of a reduced pressure; the vapor stream is then condensed, and the compounds are recovered as liquids. Contrarily, components that interact weakly with the membrane remain in the liquid stream, which facilitates coupling of pervaporation with continuous processes. Pervaporation involves not only mass transfer but also heat transfer, provided by the feed liquid stream through its enthalpy of vaporization (Anilkumar & Roy, 2024; Castro-Muñoz, 2024; Silva et al., 2021).

Pervaporation separation may be coupled with other physical and/or chemical processes to increase productivity (Kusmiyati & Susanto, 2015). For example, pervaporation can be integrated with bioreactors to provide high-quality biofuels. In the conventional process of bioethanol production, a filtered fermentation broth is subjected to distillation yielding a circa (ca.) 96 mol % ethanol solution that cannot be further purified by distillation due to the presence of an azeotrope and, consequently, cannot be used as biofuel because of the potential corrosion in the engine and/or storage tank (Yang et al., 2024); furthermore, the high energy cost of distillation impairs the economy of the process with respect to fossil fuel production (Kaewkannetra et al., 2012; Kalahal et al., 2021). Besides overcoming these disadvantages and making a more energy-efficient and competitive process, pervaporation can be implemented as a simultaneous fermentation-separation system, allowing for the removal of alcohol *in situ* and making inhibition of microorganisms during the fermentation process less likely (Jaimes et al., 2021; Van Goethem et al., 2023). Therefore, an increase in yield, productivity and economic efficiency of the overall process is expected (Shin et al., 2015; Van Goethem et al., 2023).

Alcohol concentration in the fermentation broth has been identified as a key parameter in the coupled fermentation-pervaporation process (D. Liu et al., 2015; Fu et al., 2016; Jaimes et al., 2021; Jung et al., 2018; Setlhaku et al., 2013; Wagner & Gasch, 2023; Yang et al., 2024). In particular, the concentration of the recovered alcohol (i.e., in the permeate stream) is directly proportional to the concentration of the alcohol in the broth (i.e., the feed stream). However, there is a trade-off between alcohol flux and selectivity (Z. Zhang et al., 2021) because of a possible increased diffusion of the smaller water molecules (Dong et al., 2014; L. Liu et al., 2021; Wagner & Gasch, 2023). Total flux, separation factor, and selectivity are also directly proportional to the temperature of the pervaporation cell. The increased flux is related to a higher solubility and diffusion rate of the permeable compound on the surface and bulk of the membrane, respectively (Kaewkannetra et al., 2012), due to an increased molecular mobility of the polymer chains at higher temperatures (Z. Zhang et al., 2021), which, in turn, leads to enlarging the free volume of the membrane (Dadi et al., 2018; Yang et al., 2024). For example, using PDMS membranes, Trinh et al. (2019) reported an increase in flux from 50.1 to 90.8 g/m² h, and in separation factor

from 1.9 to 2.3 as the temperature was raised from 20 to 40°C. Similarly, Lazarova et al. (2012) reported an increment in separation factor with temperature (from 25 to 75 °C) with poly(octylmethylsiloxane) (POMS) membranes. However, as the temperature approached 95°C, the separation factor showed values similar to those obtained at 50°C. On the other hand, using carbon nanotube (CNT)-loaded PDMS membranes (Xue, Wang, et al., 2016) questioned the simple relationship between alcohol flux/separation factor with temperature, arguing that weak water-ethanol hydrogen bonding interactions would not lead to the coupling between these two molecules, a phenomenon observed mainly in porous membranes. Ethanol molecules, being more voluminous, reduce the number of water molecules passing through the membrane because the interaction between ethanol and water would be weaker. Therefore, the permeability of the membrane is altered, favoring the passage of ethanol compared to water. Therefore, the differences in molecular size and hydrogen bonding interactions are key factors that explain why ethanol molecules can reduce the passage of water through the membrane (Xue, Wang, et al., 2016).

Alcohol concentration and temperature, however, are not the only factors that can be exploited to impact the flux and separation factor (or selectivity) of the membrane. The objective of this review is to highlight other variables that can be tuned to obtain high-quality fuel alcohols (bioethanol or biobutanol) through the coupled fermentation-pervaporation processes, such as the addition of some fermentation by-products or the modification of the alcohol-membrane interactions, which are paramount in the pervaporation process. Furthermore, the use of other strategies to improve fuel alcohol yield and productivity is highlighted, such as new bacterial strains, lignocellulose-based biomass to ensure food security, or the coupling of other separation processes to the fermentation-pervaporation system. Finally, several cases are presented where the coupled system produces not only the biofuel but also value-added products, making pervaporation a more attractive technique.

A comprehensive review of the scientific literature published between 2012 and 2024 was conducted. The search included keywords such as “biofuel”, “bioethanol”, or “biobutanol”, and “fermentation” and “pervaporation”. Additionally, to broaden the scope of the results, the keyword “fermentation by-products” was also included in the search. Thus, the search equation used was “biofuel” OR “bioethanol” OR “biobutanol” AND (“fermentation” OR “fermentation by-products”) AND “pervaporation”. The number of publications on biobutanol has experienced exponential growth during the analyzed period, likely due to its technical advantages over bioethanol, including a higher energy content and lower volatility. The bibliometric analysis identified three fields of intense research: 1) optimization of the fermentation process, particularly the selection of appropriate substrates and conditions to increase efficiency in obtaining biobutanol; 2) development of separation technologies, pervaporation being the most prominent between 2014 and 2023; 3) understanding of the phenomenology involved in biofuel production, whether in

fermentation or separation technologies; these key aspects have been addressed through scientific cooperation between the United States, China, and some European countries. In summary, there is a growing interest in conducting research into sustainable bioethanol and biobutanol production, with a particular focus on optimizing fermentation processes, improving separation technologies, and valorizing fermentation by-products. These advances reflect a continued commitment to biofuels development and the consolidation of a biorefinery-based approach.

Effect of Fermentation Broth Composition and Membrane Material on Pervaporation Performance

Some by-products of the fermentation process can influence pervaporation performance, impacting both the flux and separation factor. Furthermore, the addition of species, such as inorganic salts or microorganisms, can also modify flux and separation factor. Table 1 lists some species commonly present in fermentation broths and their effect on separation performance. Particularly, the presence of furfural and the addition of inorganic salts such as ammonium sulphate ($(NH_4)_2SO_4$) and calcium chloride ($CaCl_2$) have been shown to increase the flux and separation factor of alcohols, mainly in the production of bioethanol.

Membrane material is a critical factor in pervaporation performance, particularly in separation processes such as ethanol production. Polydimethylsiloxane (PDMS) is one of the most widely used membrane materials because of its high permeability to organic compounds, as well as its thermal and chemical resistance. To optimize its selectivity and performance in various applications, e.g., by reducing swelling in the presence of high ethanol concentrations and/or improving mechanical and thermal stability, bulk or surface modifications of PDMS membranes have been reported (Chen et al., 2014; Kanemoto et al., 2016). These modifications are summarized in Figure 2 and briefly discussed below.

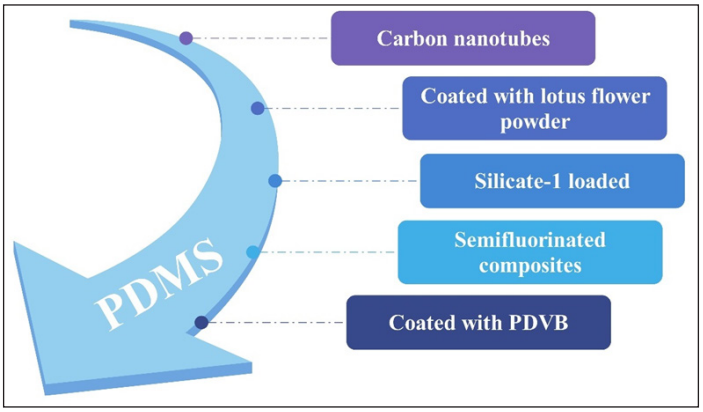


Figure 2. Polydimethylsiloxane (PDMS)-based membranes for bioethanol production
Note. PDVB = Polydivinylbenzene

Table 1
Effect of various species present in the fermentation broth on the performance of different membranes used in pervaporation

Membrane type	Species	Remarks	Reference
Silicalite-1-loaded PDMS	Furfural	Increase in ethanol flux. Furural is also permeable to PDMS membranes and can be recovered as a value-added product.	Cai et al. (2013) Chen et al. (2014) Gaykawad et al. (2013) Zhu et al. (2019)
	5-HMF	Decrease in water flux, ethanol flux not affected (ethanol production).	
	Vanilin	Decrease in ethanol flux, constant membrane selectivity (ethanol production).	
	4-hydroxybenzaldehyde	Decrease in total flux, reduction in water flux (ethanol production).	
	Catechol	Gradual decrease in total flux due to the obstruction by the oxidized form of catechol (benzoquinone).	
	Syringaldehyde	Decrease in total flux (ethanol production).	
	Acetic acid	Increase in water flux (butanol production).	
	Inorganic salts	Increase in flux, selectivity, and separation factor of bioalcohol (ethanol and butanol production) due to *salting-out*, which reduces the amount of free-water molecules and modifies the activity coefficients of the compounds in the feed.	
Silicate-1 coated with silicone rubber	Peptides (tryptone)	Decrease in total flux due to the zwitterionic nature of the peptides that reduces the hydrophobicity of the membrane.	Ikegami et al. (2014)
	Vitamins	If water-soluble, vitamins decrease selectivity and separation factor (butanol production) because of the reduction of the hydrophobicity of the membrane.	
Tubular silicate-1	Phenols	Decrease in total flux due to adsorption on porous silicate. The addition of activated carbon fixes this issue.	Kanemoto et al. (2016)
Silicate-1 loaded PDMS tubular silicate-1	Sugars	Increase in alcohol selectivity and decrease in total flux (ethanol production). Other sugars, such as glucose, increase the viscosity of the fermentation broth and resistance to ethanol transport, reducing its flux and partition factor. Sugars can also block the solvent diffusion pathways (e.g., zeolite channels).	Chen et al. (2014) Kanemoto et al. (2016)

Table 1 *(continue)*

Membrane type	Species	Remarks	Reference
PDMS Ceramics hollow fiber supported PDMS PET supported PDMS	Microorganisms	Effect depends on the membrane material: in silicate-loaded membranes, increase in alcohol flux with that of other components remaining constant by the addition of yeast (ethanol production); in PET (polyethylene terephthalate)-supported PDMS membranes, decrease in flux due to clogging of the membrane without effect on other membrane parameters.	Cai et al. (2016) Chen et al. (2014) Dong et al. (2014) Jaimes et al. (2021) Jung et al. (2018) Kaewkannetra et al. (2012)

Note. PDMS = Polydimethylsiloxane; 5-HMF = 5-hydroxymethylfurfural; PET = Polyethylene terephthalate

Incorporation of inorganic fillers such as nanoparticles or zeolites/silica into the PDMS matrix has been used to enhance the selective separation of larger organic molecules while preventing the transport of water. For instance, the addition of silicalite-1 into PDMS membranes improves the ethanol separation factor and reduces the flux (Chen et al., 2014; Santos et al., 2018) due to the hydrophobicity of silicalite and its preferential retention of ethanol, which prevents water permeation. However, above a certain threshold (saturation load), the ethanol flux decreases because of a reduction in the free volume of the membrane (Santos et al., 2018). Furthermore, silicalite-1-PDMS membranes would not be suitable for the separation of acetone, butanol, and ethanol (ABE) fermentation broth (especially butanol) because the acetic acid (fermentation by-product) decreases the selectivity by increasing water flux, as reported in Table 1 (Chen et al., 2014). It has also been documented that the incorporation of 10% CNTs in PDMS membranes increases the ethanol flux and the separation factor by 45.7 and 33.8, respectively (Xue, Wang, et al., 2016).

Other nanoscale effects have been exploited, such as the lotus effect, i.e., the ability of certain materials to repel water, analogous to how lotus leaves repel water due to their nanostructured surface (Xue, Wang, et al., 2016). For example, He et al. (2020) mixed PDMS with lotus flower powder to improve hydrophobicity, but the relatively large particles used created voids and defects in the membrane matrix, which compromised selectivity by allowing unwanted components to pass through, lessening membrane performance. Despite the drop in selectivity, the flux of the lotus flower powder-coated membrane increased by approximately 25% relative to the PDMS membrane, particularly when approximately 1.5 wt.% of lotus-flower powder was used.

Li et al. (2014) reported the use of a PDMS and polyacrylonitrile (PAN) membrane loaded with silicalitate-1 for the recovery of butanol from cassava-derived fermentation broths. The membrane achieved a separation factor between 29 and 32, with butanol flux ranging from 118.70 to 203.76 g/m^2 h. These results, comparable to those obtained with simulated models, suggest that the membrane did not experience fouling when used with the fermentation broth. Similarly to ethanol recovery, carbon nanotube modified membranes have also been used for butanol separation (Xue, Liu, et al., 2016).

F. Wang et al. (2024) employed slot coating to form vinyl-PDMS membranes on a porous support in as little as 20 min, achieving selective layers with tunable thicknesses from 0.4 to 15 μm. The resulting membrane exhibited a total flux of 2.8 kg/m² h and a separation factor of 10.3 when tested with a 5 wt.% ethanol/water mixture at 60°C. Furthermore, the authors reported that the incorporation of nanoparticles improved the permeability by up to 35% and increased the separation factor by approximately 15, allowing the membrane properties to be fine-tuned to specific process requirements.

On the other hand, the issue of compatibility between organic membranes and inorganic silicalite-1 particles needs to be considered, lest it negatively impacts the membrane

selectivity. To this end, membranes incorporating metal-organic frameworks (MOFs), such as zeolitic imidazolate frameworks (ZIFs), have been proposed (S. Liu et al., 2013). The addition of this material in PDMS membranes improved the hydrophobicity and, therefore, increased the butanol separation factor, but organic acids present in the fermentation broth cause non-selective defects, such as interstices, which may significantly reduce the butanol separation factor and increase the flux. Si et al. (2019) reported that to avoid degradation, PDMS membranes were loaded with ZIF-derived carbon nanopores, which prevent membrane deterioration by organic acids and allow the membrane performance to match that observed with simulated models.

Chemical modification of the PDMS surface by introduction of functional groups has also been reported to augment the affinity for specific compounds, thus increasing the selectivity of the separation (Ikegami et al., 2014). Furthermore, blending PDMS with other polymers results in hybrid membranes with improved mechanical or selective properties according to the process requirements. Cross-linking, which involves the interconnection of PDMS chains, improves stability and resistance to swelling, allowing the membrane to perform better at higher ethanol concentrations (Chen et al., 2014).

Polydivinylbenzene (PDVB) coating has shown a 13% increase in the separation factor and a 30% increase in flux compared to the uncoated membrane. In addition, PDVB prevented the membrane from wetting and accelerated ethanol uptake. Semifluorinated self-assembled monolayer membranes with high hydrophobicity, prepared by ultraviolet (UV)/ozone photooxidation, have also been described. The fabrication conditions of membranes significantly influence their performance during pervaporation. Initially, a membrane operating at 60°C presented a high flux (412.9 g/m² h), but the separation factor was 13.1, suggesting a limited ability to effectively discriminate between the mixture components (ethanol and water). The membrane selectivity peaked with time-on-stream, with separation factors of 11.8, 13.1, and 9.7 at 30, 60, and 90 min, respectively, pointing to changes in the membrane structure or properties during the process (Li et al., 2013).

On the other hand, the properties of styrene-butadiene-styrene (SBS) membranes can be modified in the presence of a PDMS or polyurethane (PU) membrane (Santoro et al., 2017). When the SBS membrane is combined with a PU membrane, the flux increases due to the hydrophobicity of PU, while the selectivity improves when combined with a PDMS membrane. Both types of membranes showed increased flux as the feed ethanol concentration increased (from 3 to 9 wt.%), although their separation factor decreased. Optimal performance of these membranes was also achieved at 50°C and low ethanol concentrations (3 wt.%) to improve selectivity, or at high ethanol concentrations to promote flux (Santoro et al., 2017). In addition, the embodiment of self-assembled polystyrene block copolymer into PDMS membranes has attracted attention for its ability to improve membrane flux (increasing from 110 to 220 g/m^2 h), while achieving higher separation

factors for acetone (7-fold increase) and butanol (3-fold increase) from a glucose-derived ABE fermentation broth (Shin et al., 2015).

Other materials different from PDMS have also been evaluated. Among them, polyvinyl alcohol (PVA) membranes coated with cellulose acetate are not attractive for implementation in the separation of alcohols from fermentation broths since the yields are reduced by up to 50% compared to simple aqueous solutions of 14.6% ethanol (Kaewkannetra et al., 2012). Polymer of mixed solvents (POMS) also showed poor performance, i.e., a permeate with an ethanol concentration between 1.5-1.8 wt.% when the concentration in the feed stream varied between 0.5-2.5 wt.% (Z. Zhang et al., 2021). POMS has also been used in separation membranes for butanol. Compared to PDMS membranes, a higher flux was achieved by using a fermentation broth derived from wheat straw, despite operating under conditions that led to a lower driving force; thus, for the POMS membrane, a pressure of 20 mbar and a temperature of 32°C were used, in comparison with 9-14 mbar and 37°C with PDMS. Similarly, the separation factor of the POMS membrane was higher than that of the PDMS membrane (Van Hecke & De Wever, 2017). Another polymer used for butanol separation in systems coupled to liquid-liquid extraction is sulfonated poly(arylene ether sulfone) (SPAES). This membrane provides a protective effect on the fermenting microorganisms against the solvent used in the extraction. Furthermore, it is noted that SPAES exhibited an intermediate affinity compared to membranes made with polyamide (PA) and PDMS (Kim et al., 2020). On the other hand, membranes made from polyetherimide (PEI) loaded with graphene oxide have been designed. The presence of epoxy, carboxyl, or hydroxyl functional groups of graphene oxide enhanced the diffusion of the feed mixture. PEI demonstrated a separation factor of 99.3, suggesting it as a promising membrane for butanol dehydration (Manshad et al., 2020).

The use of catalytically active membranes, which can immobilize the microorganisms responsible for fermentation, has also been reported. These membranes exhibited similar flux and separation factors to PDMS membranes (Cao et al., 2020). Promising results have been obtained with hydrophilic membranes specifically designed to remove water while enriching ethanol in the feed stream. However, in a simultaneous fermentation and pervaporation process, this approach could be problematic, as rapid water removal can hamper microbial activity essential for fermentation. Microorganisms, such as *Saccharomyces cerevisiae*, require water for growth and metabolism, and excessive water removal could limit ethanol production. Furthermore, high ethanol concentrations could become toxic to microorganisms, reducing fermentation yield, making it important to balance the ethanol concentration to achieve maximum process efficiency. Some of these membranes are made from PVA and graphene, where the latter creates nanochannels within the membrane matrix that host water monolayers. Increasing the loading of graphene improved the flux and selectivity towards water, while reducing the ethanol flux to almost

zero (Kalahal et al., 2021). Similarly, high water permeability carbon molecular sieve membranes (CMSM) have been described, as well as other types of hydrophilic membranes developed by Rahimalimamaghani et al (2022). From these works, it was concluded that to attain a high ethanol concentration, membranes should be designed to achieve separation factors close to 10,000 when operating at 5% ethanol concentration in the feed stream (Nagy et al., 2015).

The incorporation of alkyl fluorinated groups has been reported to mask the clogging caused by the microorganisms in the fermentation broth, yielding a 37% increase in flux, a 104% increase in separation factor, and high stability during continuous butanol removal compared to membranes without alkyl fluorinated groups (Zhu et al., 2020).

Polyether-block-amide (PEBA) loaded with ZIF-71 metal-organic framework membranes has been reported for butanol pervaporation. The increased hydrophobicity of this structured network enhances the affinity of the membrane for butanol, influencing both the flux and separation factor, according to S. Liu et al. (2013), membrane flux and separation factor are maximized at 20 and 25% ZIF-71 (Figure 3) loadings, respectively, while a reduction in flux and separation factor was observed for acetone and ethanol present in the ABE fermentation broth.

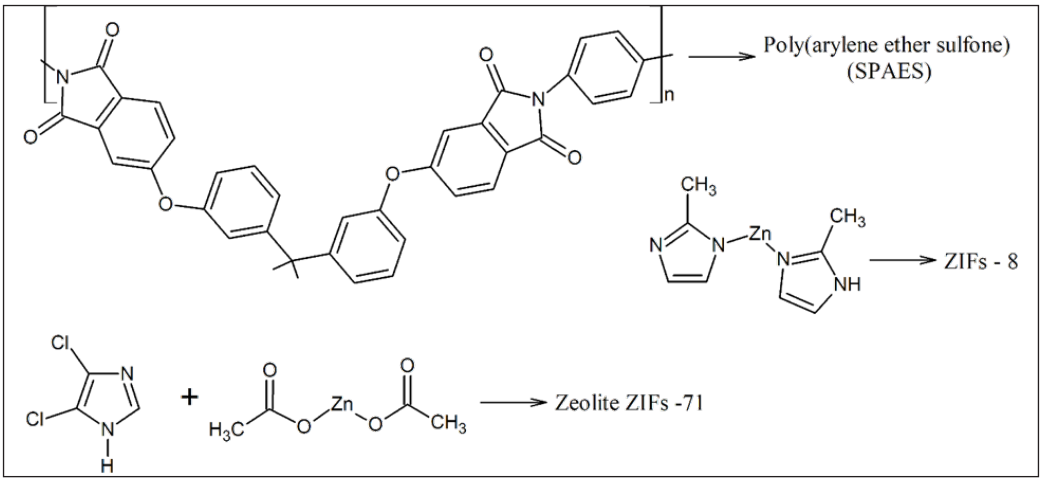


Figure 3. Membranes reported in biobutanol production
Note. SPAES = Poly(arylene ether sulfone); ZIF = Zeolitic imidazolate framework

Komal et al. (2024) investigated the need for efficient approaches to traditional butanol separation methods, which are costly and energy-intensive. In their study, a novel strategy was implemented using mixed-matrix membranes (MMMs) incorporating ZIF-8 in liquid form (ZIF-8-PL) into PDMS. Compared to conventional membranes (unmodified PDMS), membranes containing ZIF-8-PL showed significant performance improvement, achieving

a ca. 102% increase in flux and a 15% improvement in separation factor, highlighting the advantages of ZIF-8. Therefore, MMMs with liquid fillings (PL) represent a promising approach to optimize energy efficiency and reduce costs in industrial-scale butanol separation.

Several ionic liquids have been proposed in combination with PVA and polytetrafluoroethylene (PTFE) membranes (Figure 4). It has been established that the presence of these ionic liquids increases the flux, particularly with 1,3,5-Tris(bromopropyl) benzene (TBP-Br), due to the high hydrophobicity and flexibility of the polymer chains, which expands the free volume of the membrane. However, the separation factor increased with higher loadings of 1,3,5-Tris(bromomethyl)-2,4,6-trimethylbenzene chloride (TBRDP-Cl) or 1,3,5-Tris(bromomethyl)benzene (TBMP), reaching a maximum value of 147 for the membrane with 30% 1,3,5-tris(bromomethyl)phenyl-functionalized mesoporous silica (TBMP-MS), while the opposite effect was observed with TBP-Br (Sajjad et al., 2019). Contrarily, it is necessary to use a gelling agent to prevent the leaching of other ionic liquids out of the membrane when working at low pressures, thus ensuring mechanical stability during operation, as was reported by Plaza et al. (2013) with the ionic liquid 1-Butyl-3-methylimidazolium hexafluorophosphate (BMIM-PF6) on a polytetrafluoroethylene (PTFE) membrane: the interaction with the membrane led to preferential transport of butanol, which was the cause of the higher partition coefficient.

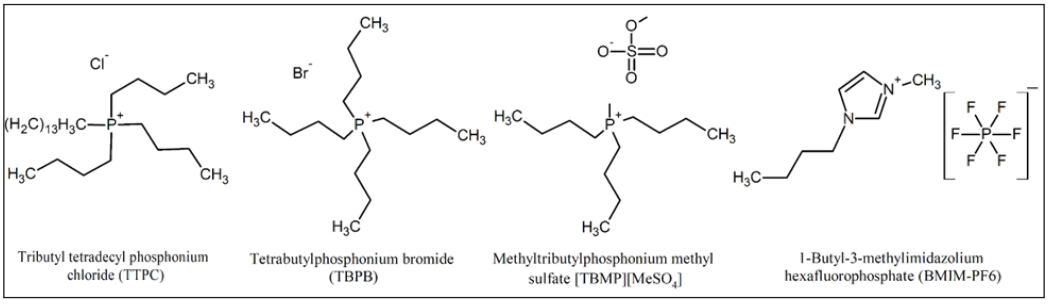


Figure 4. Ionic liquids used in polyvinyl alcohol and polytetrafluoroethylene membranes for butanol pervaporation

Approaches to Improve Pervaporation Performance

Depending on the type of fuel alcohol, a wide variety of bacterial strains and biomass have been studied for the fermentation processes, as shown in Figure 5. Bioethanol production has been reported from various sources, mainly those with high polysaccharide content that do not endanger food safety, such as wastes of banana (Wagner & Gasch, 2023), corn (Chen et al., 2014), newspaper (Trinh et al., 2013), coffee (Nguyen et al., 2017), or cruciferous vegetables (Song, Nguyen, et al., 2017), as well as non-waste sources such as Jerusalem artichoke (Song, Oh, et al., 2017), bamboo (Song et al., 2020), rye straw, and willow wood

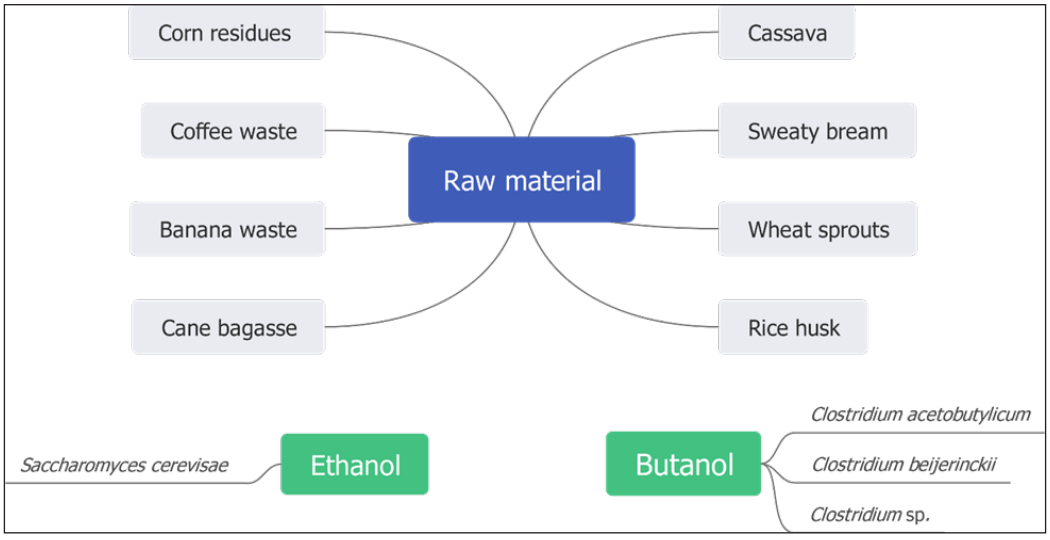


Figure 5. Sources and microorganisms are available to produce ethanol and butanol by pervaporation coupled with fermentation

chips (Gaykawad et al., 2013). The fermentation is mainly carried out by the *S. cerevisiae*. However, the use of other types of yeast has also been described, such as autoflocculants resulting from the fusion of protoplasts of *S. cerevisiae* and *Schizosaccharomyces pombe* strains (Xue, Wang, et al., 2016). The reduction of the complex sugars from these sources, required to carry out the fermentation and achieved by a previous stage of acid hydrolysis, is key to improving fermentation performance since microorganisms such as *Clostridium pasteurianum* or *S. cerevisiae* mainly metabolize simple sugars, such as glucose. Therefore, this stage must be carefully controlled to balance the higher amount of sugars obtained at higher acid concentrations with the concomitant formation of byproducts that could affect membrane performance (de Andrade et al., 2021).

It has also been proposed to use ionic liquids such as 1-butyl-3-methylimidazolium chloride combined with inorganic acids to dose the acid during hydrolysis (Trinh et al., 2019), increasing the production of fermentable sugars and, as a result, a higher ethanol yield in the fermentation broth and a higher ethanol concentration in the permeate. Additionally, yeast immobilization has been evaluated to improve ethanol productivity. For instance, Santos et al. (2018) reported that the use of calcium alginate to immobilize *S. cerevisiae* resulted in a 1.4-fold increase in ethanol concentration.

From a land use and agricultural efficiency perspective, bioethanol is a better option than biodiesel, as approximately 420 gallons/acre of ethanol can be produced compared to only 60 gallons/acre of biodiesel from soybeans. Furthermore, depending on the feedstock and process, soybean oil prices could increase with increased biodiesel production.

However, technical hurdles, such as corrosiveness, water miscibility, low energy density, and volatility losses, complex higher bioethanol blends. In contrast, *n*-butanol is a superior alternative fuel due to its higher heating value, and lower corrosiveness, polarity, volatility, and heat of vaporization, which minimize ignition problems in internal combustion engines. Furthermore, blends of butanol and petroleum-derived fuels may be used without engine modification. In addition, butanol produces lower carbon dioxide (CO_2), nitrogen oxide (NO_x), and sulphur oxides (SO_x) emissions compared to ethanol or gasoline. These advantages, coupled with the significant global demand for biofuels (approximately 51 billion gallons in 2022), are driving increased interest in the sustainable production of biomass and organic wastes, thereby reducing dependence on fossil fuels (Kanemoto et al., 2016; Lee & Lin, 2024).

Biobutanol production has been reported using a variety of bacterial strains to ferment materials derived from sweet sorghum (Cai et al., 2013; Kanemoto et al., 2016), wheat, sugarcane bagasse (Van Hecke & De Wever, 2017), and corn (C. Zhang et al., 2019, 2021) (Figure 5). Cassava has also been considered as a potential source of biobutanol due to its ability to grow in a variety of soils, thus avoiding competition with other crops (Li et al., 2014). Among the innovations in butanol production, the use of bacterial strains other than *Clostridium acetobutylicum* has been reported. Jung et al. (2018) documented the use of Enterobacter aerogenes to produce isobutanol, a biofuel alternative. In particular, about 83% of the isobutanol was recovered at concentrations ranging from 55 to 226 g/L using a PDMS membrane (Jung et al., 2018). The ability of *Clostridium* sp., on the other hand, to produce fuel-grade butanol without acetone as a by-product is very promising because it eliminates the need for an additional step to separate the butanol-acetone mixture (all membranes reported to date exhibit high affinity for both acetone and butanol) (Xin et al., 2016).

A feasibility study for obtaining butanol with *C. pasteurianum* GL11 demonstrates remarkable progress in the sustainable and cost-effective production of biofuels, making this strain an excellent candidate for industrial applications. In batch fermentation using glycerol, *C. pasteurianum* GL11 achieved a butanol concentration of 14.7 g/L with a yield of 0.41 g/g, while removing by-products such as acetone and 1,3-propanediol (1,3-PDO), which simplifies separation processes and improves productivity. Furthermore, when *in situ* extraction was applied during fed-batch fermentation, butanol reached 28.8 g/L, demonstrating the potential of integrating advanced extraction techniques to scale up production. Another key feature of *C. pasteurianum* GL11 is its ability to hydrolyze complex polysaccharides such as starch and xylan into butanol. This process is facilitated by the secretion of xylanases and amylases, allowing for consolidated bioprocessing that reduces the need for external enzymatic treatments and lowers substrate preparation costs. This combination of efficient substrate use, high productivity, and cost reduction positions

C. pasteurianum GL11 as a promising candidate for industrial-scale butanol production (Jung et al., 2018; Xin et al., 2016). Indeed, its ability to utilize low-cost substrates, such as glycerol (a biodiesel byproduct) and polysaccharides, as well as its high tolerance to butanol and optimized metabolic profile, reduces the energy demands of separation techniques, making the process economically viable and environmentally sustainable. Future research may focus on optimizing fermentation conditions and process scalability. Furthermore, co-fermentation with other renewable substrates can further diversify applications and improve productivity. This strain highlights the potential of combining genetic traits and innovative bioprocessing strategies to address global energy challenges while efficiently utilizing industrial waste streams (Xin et al., 2016).

Su et al. (2024)'s co-fermentation using dough and okara waste with *Clostridium* sp. strain BOH3 showed notable improvements in biobutanol production compared to individual fermentation. Using 80 g/L of a co-substrate composed of 85% mash and 15% okara, 15.9 ± 0.90 g/L of butanol and 25.4 ± 1.43 g/L of total solvents (acetone-butanol-ethanol) were achieved. These results represent increases of 91.6% in butanol and 108.2% in total solvents compared to the mash-only fermentation. The okara-only fermentation resulted in minimal solvent production due to its unfavorable carbon-to-nitrogen ratio. Furthermore, cofermentation accelerated 1.7-fold the transition from acidogenesis to solventogenesis and increased enzyme activity (amylase and butanol dehydrogenase). These findings highlight cofermentation as an effective strategy to optimize the conversion of food waste into biobutanol sustainably.

Fermentation-pervaporation systems can operate under batch, fed-batch, and continuous conditions. Although the mode of operation does not directly influence the butanol concentration in the permeate, since the mass transfer mechanism through the membrane is not modified (unless the membrane itself is changed) (Cai et al., 2016; Fu et al., 2016; H. Zhang et al., 2023; L. Liu et al., 2021; Li et al., 2014; Shin et al., 2015) continuous removal of butanol ensures a higher fuel yield due to the mitigation of butanol-induced bacterial toxicity. Fractional condensation coupled to vapor compression and PDMS membranes has been explored to replace conventional condensation at low temperatures, resulting in the production of high-concentration ethanol (50%) with a selectivity of 20 and energy savings of up to 50% (Z. Wang et al., 2023). Stripping has also been employed as a pretreatment operation prior to pervaporation to remove butanol and other substances that could affect the membrane flux due to fouling, achieving high fuel concentrations, i.e., 44.7 g/L (Xue, Liu, et al., 2016) and 55.9 g/L (Rochón et al., 2020).

Challenges and Opportunities

In addition to the production of biofuels, the simultaneous fermentation-pervaporation process has also been used for the production of other valuable products. Mannose (a sugar)

can be obtained from lignocellulosic biomass such as coffee grounds, reaching a purity of 95.3% (Nguyen et al., 2017). Psicose, an epimer of fructose, is obtained from inulin present in cruciferous plant wastes (Song, Nguyen, et al., 2017) and Jerusalem artichoke (Song, Oh, et al., 2017), and which has the potential to suppress hepatic lipogenic enzymes, has been produced together with ethanol using the strain *S. cerevisiae*. Xylose, which can be made from the fermentation of bamboo pretreated with hydrogen peroxide and acetic acid in the presence of *S. cerevisiae* and that, like psicose, cannot be metabolized by this yeast to ethanol, was obtained with high yield (89.6%, retentate stream) along with ethanol (82.6%, permeate stream) using PVA membranes. Furthermore, given the high availability of carbohydrates, this strategy could reduce water consumption up to 2.5 times per kilogram of ethanol obtained (Van Hecke et al., 2016).

Other approaches to reduce effluents have been reported by C. Zhang et al. (2019, 2021), who used the retentate as a buffer and nutrient source for the fermentation process, achieving savings of 86 to 92% in the feed stream and recovering 94% of ABE from the fermentation of corn residues with *C. acetobutylicum*. Despite the presence of acids that can be toxic to microorganisms during fermentation, no significant deleterious effect was observed because of the conversion of acids into lipids by using *Rhodotorula glutinis*.

While Outram et al. reported improvements in fermentation yield of up to 67% (Outram et al., 2017), Xue, Wang, et al. (2016) reached ethanol concentrations of 615.1 g/L in the permeate with CNT/PDMS membranes. Furthermore, using a zeolite-PDMS membrane, they achieved 253 g ABE/L in the condensate (Xue et al., 2015). This represents an increase in ethanol concentration from 3 to 4% compared to conventional fermentations, with a maximum ethanol purity in the permeate of 78%. Similarly, in ABE fermentation combined with pervaporation, ethanol purity in the permeate ranges from 2 to 27%. These findings suggest that hybrid membranes can significantly reduce energy consumption in the recovery of biobutanol and bioethanol. Currently, pervaporation membranes are limited to laboratory-scale development, facing challenges in scalability, cost optimization, and integration with fermentation processes. Key issues include improving durability, stability, and long-term performance. Overcoming these challenges is essential to establishing pervaporation as an efficient, stable, and cost-effective solution for biobutanol and bioethanol production.

CONCLUSION AND FUTURE PERSPECTIVES

Pervaporation separation is a highly efficient operation for the recovery of organic solvents, particularly alcohols derived from fermentation processes. The integration of pervaporation and fermentation creates a system of great interest to produce biofuels, mainly bioethanol and biobutanol. To improve pervaporation performance and achieve high biofuel yield and purity, it is common practice to tune both the alcohol concentration in the fermentation broth (liquid feed stream) and the operating temperature of the pervaporation process. While

these are the most common factors evaluated, they are not the only ones that influence the overall performance of the pervaporation-fermentation system.

The presence of furfural from biomass fermentation or added intentionally, as well as the addition of inorganic salts to the fermentation broth, contributes to improving flux and selectivity. The type of membrane material can also influence the performance of the fermentation-pervaporation system. The performance of PDMS membranes can be further improved by incorporating functional groups (amino, carboxyl, or sulfonyl) or species such as glycerol and sorbitol. In addition, loading of nanomaterials such as graphene, silicon nanocomposites, or carbon nanotubes can also improve the separation.

It is necessary to devise alternative polymers to PDMS that present a high affinity for alcohols, which can be achieved by establishing the chemical structure that optimizes the interaction of the alcohol with the membrane under specific conditions. Increasing the concentration of fermentable sugars during biomass pretreatment (hydrolysis) is another variable that inherently improves pervaporation performance, as it leads to higher biofuel concentrations in the fermentation broth.

From a different perspective, the use of strains, such as *Zymomonas mobilis* or *Bacillus subtilis*, can be advantageous as it allows the production of the desired biofuel while minimizing the generation of other solvents, such as acetic acid, butyric acid, ethers, and aldehydes. In addition to fermentation, integrating operations such as distillation, extraction, adsorption, crystallization, filtration, and ion exchange with pervaporation can optimize process performance. These ancillary techniques not only improve product purity and recovery but can also achieve energy savings of up to 50%. By combining these methodologies, the overall efficiency of biofuel production would be maximized.

In summary, pervaporation has gained increasing interest and importance as a separation method for biofuel production due to its dual capacity to produce value-added sugars and reduce the amount of raw material required for large-scale biofuel production. Therefore, research into new materials, genetic modifications of microorganisms, and development and implementation of additional operational processes are key points in the field of biofuel production by fermentation and pervaporation.

ACKNOWLEDGEMENTS

The authors thank the Pedagogical and Technological University of Colombia for their financial support of this work through project DIN SGI 3491. Additionally, Jorge Moreno-Cascante wishes to express his gratitude to the Ministry of Science and Technology (Minciencias) for the financial support to pursue his doctorate through the call "Formación de capital humano de alto nivel – Doctorado, en el marco de la Segunda cohorte de convocatoria N°1 del plan bienal (2018–2020)."

REFERENCES

Anilkumar, A., & Roy, Y. (2024). Improved solution thermodynamics modeling in pervaporation. *Chemical Engineering Science*, *299*, 120497. https://doi.org/10.1016/j.ces.2024.120497

Cai, D., Li, P., Chen, C., Wang, Y., Hu, S., Cui, C., Qin, P., & Tan, T. (2016). Effect of chemical pretreatments on corn stalk bagasse as immobilizing carrier of *Clostridium acetobutylicum* in the performance of a fermentation-pervaporation coupled system. *Bioresource Technology*, *220*, 68–75. https://doi.org/10.1016/j.biortech.2016.08.049

Cai, D., Zhang, T., Zheng, J., Chang, Z., Wang, Z., Qin, P.-Y., & Tan, T.-W. (2013). Biobutanol from sweet sorghum bagasse hydrolysate by a hybrid pervaporation process. *Bioresource Technology*, *145*, 97–102. https://doi.org/10.1016/j.biortech.2013.02.094

Cao, Z., Xia, C., Jia, W., Qing, W., & Zhang, W. (2020). Enhancing bioethanol productivity by a yeast-immobilized catalytically active membrane in a fermentation-pervaporation coupling process. *Journal of Membrane Science*, *595*, 117485. https://doi.org/10.1016/j.memsci.2019.117485

Castro-Muñoz, R. (2024). Composite 2D material-based pervaporation membranes for liquid separation: A review. *Molecules*, *29*(12), 2829. https://doi.org/10.3390/molecules29122829

Chen, J., Zhang, H., Wei, P., Zhang, L., & Huang, H. (2014). Pervaporation behavior and integrated process for concentrating lignocellulosic ethanol through polydimethylsiloxane (PDMS) membrane. *Bioprocess and Biosystems Engineering*, *37*, 183–191. https://doi.org/10.1007/s00449-013-0984-5

Dadi, D., Beyene, A., Simoens, K., Soares, J., Demeke, M. M., Thevelein, J. M., Bernaerts, K., Luis, P., & Van der Bruggen, B. (2018). Valorization of coffee byproducts for bioethanol production using lignocellulosic yeast fermentation and pervaporation. *International Journal of Environmental Science and Technology*, *15*, 821–832. https://doi.org/10.1007/s13762-017-1440-x

de Andrade, C. M., Cogo, A. J. D., Perez, V. H., dos Santos, N. F., Okorokova-Façanha, A. L., Justo, O. R., & Façanha, A. R. (2021). Increases of bioethanol productivity by *S. cerevisiae* in unconventional bioreactor under ELF-magnetic field: New advances in the biophysical mechanism elucidation on yeasts. *Renewable Energy*, *169*, 836–842. https://doi.org/10.1016/j.renene.2021.01.074

Dong, Z., Liu, G., Liu, S., Liu, Z., & Jin, W. (2014). High performance ceramic hollow fiber supported PDMS composite pervaporation membrane for bio-butanol recovery. *Journal of Membrane Science*, *450*, 38–47. https://doi.org/10.1016/j.memsci.2013.08.039

Fu, C., Cai, D., Hu, S., Miao, Q., Wang, Y., Qin, P., Wang, Z., & Tan, T. (2016). Ethanol fermentation integrated with PDMS composite membrane: An effective process. *Bioresource Technology*, *200*, 648–657. https://doi.org/10.1016/j.biortech.2015.09.117

Gaykawad, S. S., Zha, Y., Punt, P. J., van Groenestijn, J. W., van der Wielen, L. A. M., & Straathof, A. J. J. (2013). Pervaporation of ethanol from lignocellulosic fermentation broth. *Bioresource Technology*, *129*, 469–476. https://doi.org/10.1016/j.biortech.2012.11.104

He, X., Wang, T., Huang, J., Chen, J., & Li, J. (2020). Fabrication and characterization of superhydrophobic PDMS composite membranes for efficient ethanol recovery via pervaporation. *Separation and Purification Technology*, *241*, 116675. https://doi.org/10.1016/j.seppur.2020.116675

Ikegami, T., Negishi, H., Nakayama, S., Kobayashi, G., & Sakaki, K. (2014). Pervaporative concentration of biobutanol from ABE fermentation broths by *Clostridium saccharoperbutylacetonicum* using silicone rubber-coated silicalite-1 membranes. *Separation and Purification Technology*, *132*, 206–212. https://doi.org/10.1016/j.seppur.2014.05.030

Imad, M., & Castro-Muñoz, R. (2023). Ongoing progress on pervaporation membranes for ethanol separation. *Membranes*, *13*(10), 848. https://doi.org/10.3390/membranes13100848

Jaimes, J. H. B., Alvarez, M. E. T., de Moraes, E. B., Maciel, M. R. W., & Filho, R. M. (2021). Separation and semi-empiric modeling of ethanol–water solutions by pervaporation using PDMS membrane. *Polymers*, *13*(1), 93. https://doi.org/10.3390/polym13010093

Jung, H.-M., Lee, J. Y., Lee, J.-H., & Oh, M.-K. (2018). Improved production of isobutanol in pervaporation-coupled bioreactor using sugarcane bagasse hydrolysate in engineered *Enterobacter aerogenes*. *Bioresource Technology*, *259*, 373–380. https://doi.org/10.1016/j.biortech.2018.03.081

Kaewkannetra, P., Chutinate, N., Moonamart, S., Kamsan, T., & Chiu, T. Y. (2012). Experimental study and cost evaluation for ethanol separation from fermentation broth using pervaporation. *Desalination and Water Treatment*, *41*(1–3), 88–94. https://doi.org/10.1080/19443994.2012.664682

Kalahal, P. B., Kulkarni, A. S., Sajjan, A. M., Khan, T. M. Y., Badruddin, I. A., Kamangar, S., Banapurmath, N. R., Ayachit, N. H., Naik, M. L., & Marakatti, V. S. (2021). Fabrication and physicochemical study of B2SA-grafted poly(vinyl alcohol)–graphene hybrid membranes for dehydration of bioethanol by pervaporation. *Membranes*, *11*(2), 110. https://doi.org/10.3390/membranes11020110

Kanemoto, M., Negishi, H., Sakaki, K., Ikegami, T., Chohnan, S., Nitta, Y., Kurusu, Y., & Ohta, H. (2016). Efficient butanol recovery from acetone–butanol–ethanol fermentation cultures grown on sweet sorghum juice by pervaporation using silicalite-1 membrane. *Journal of Bioscience and Bioengineering*, *121*(6), 697–700. https://doi.org/10.1016/j.jbiosc.2015.11.008

Kim, J. H., Cook, M., Peeva, L., Yeo, J., Bolton, L. W., Lee, Y. M., & Livingston, A. G. (2020). Low energy intensity production of fuel-grade bio-butanol enabled by membrane-based extraction. *Energy and Environmental Science*, *13*(12), 4862–4871. https://doi.org/10.1039/D0EE02927K

Komal, A., Kasuri, A., Yasin, M., Khan, A. L., Razzaq, A., Alhmohamadi, H., Akhtar, F. H., Raja, A. A., Nawaz, R., Abdulmalek, E., & Gilani, M. A. (2024). Enhancing butanol separation efficiency in pervaporation with ZIF-8 porous liquid infused mixed matrix membranes. *Chemical Engineering Research and Design*, *206*, 43–53. https://doi.org/10.1016/j.cherd.2024.04.044

Kusmiyati, & Susanto, H. (2015). Fuel grade bioethanol production from iles-iles (*Amorphophalus campanulatus*) tuber. *Procedia Environmental Sciences*, *23*, 199–206. https://doi.org/10.1016/j.proenv.2015.01.031

Lazarova, M., Bösch, P., & Friedl, A. (2012). POMS membrane for selective separation of ethanol from dilute alcohol-aqueous solutions by pervaporation. *Separation Science and Technology*, *47*(12), 1709–1714. https://doi.org/10.1080/01496395.2012.658943

Lee, J., & Lin, K.-Y. A. (2024). Bio-butanol production on heterogeneous catalysts: A review. *Journal of the Taiwan Institute of Chemical Engineers*, *157*, 105421. https://doi.org/10.1016/j.jtice.2024.105421

Li, J., Chen, X., Qi, B., Luo, J., Zhang, Y., Su, Y., & Wan, Y. (2014). Efficient production of acetone–butanol–ethanol (ABE) from cassava by a fermentation–pervaporation coupled process. *Bioresource Technology*, *169*, 251–257. https://doi.org/10.1016/j.biortech.2014.06.102

Li, J., Ji, S., Zhang, G., & Guo, H. (2013). Surface-modification of poly(dimethylsiloxane) membrane with self-assembled monolayers for alcohol permselective pervaporation. *Langmuir*, *29*(25), 8093–8102. https://doi.org/10.1021/la400930y

Liu, D., Liu, G., Meng, L., Dong, Z., Huang, K., & Jin, W. (2015). Hollow fiber modules with ceramic-supported PDMS composite membranes for pervaporation recovery of bio-butanol. *Separation and Purification Technology*, *146*, 24–32. https://doi.org/10.1016/j.seppur.2015.03.029

Liu, L., Wang, Y., Wang, N., Chen, X., Li, B., Shi, J., & Li, X. (2021). Process optimization of acetone-butanol-ethanol fermentation integrated with pervaporation for enhanced butanol production. *Biochemical Engineering Journal*, *173*, 108070. https://doi.org/10.1016/j.bej.2021.108070

Liu, S., Liu, G., Zhao, X., & Jin, W. (2013). Hydrophobic-ZIF-71 filled PEBA mixed matrix membranes for recovery of biobutanol via pervaporation. *Journal of Membrane Science*, *446*, 181–188. https://doi.org/10.1016/j.memsci.2013.06.025

Manshad, S., Isloor, A. M., Mohd Nawawi, M. G., Inamuddin., Khan, I., & Marwani, H. M. (2020). Pervaporation dehydration of bio-fuel (*n*-butanol) by dry thermal treatment membrane. In *Materials Research Express* (Vol. 7, No. 6, p. 065001). IOP Publishing. https://doi.org/10.1088/2053-1591/ab9562

Nagy, E., Mizsey, P., Hancsók, J., Boldyryev, S., & Varbanov, P. (2015). Analysis of energy saving by combination of distillation and pervaporation for biofuel production. *Chemical Engineering and Processing: Process Intensification*, *98*, 86–94. https://doi.org/10.1016/j.cep.2015.10.010

Nguyen, Q. A., Cho, E., Trinh, L. T. P., Jeong, J.-S., & Bae, H.-J. (2017). Development of an integrated process to produce D-mannose and bioethanol from coffee residue waste. *Bioresource Technology*, *244*(Part 1), 1039–1048. https://doi.org/10.1016/j.biortech.2017.07.169

Outram, V., Lalander, C.-A., Lee, J. G. M., Davies, E. T., & Harvey, A. P. (2017). Applied *in situ* product recovery in ABE fermentation. *Biotechnology Progress*, *33*(3), 563–579. https://doi.org/10.1002/btpr.2446

Plaza, A., Merlet, G., Hasanoglu, A., Isaacs, M., Sanchez, J., & Romero, J. (2013). Separation of butanol from ABE mixtures by sweep gas pervaporation using a supported gelled ionic liquid membrane: Analysis of transport phenomena and selectivity. *Journal of Membrane Science*, *444*, 201–212. https://doi.org/10.1016/j.memsci.2013.04.034

Rahimalimamaghani, A., Pacheco Tanaka, D. A., Llosa Tanco, M. A., Neira D'Angelo, F., & Gallucci, F. (2022). New hydrophilic carbon molecular sieve membranes for bioethanol dehydration via pervaporation. *Chemical Engineering Journal*, *435*(Part 1), 134891. https://doi.org/10.1016/j.cej.2022.134891

Rochón, E., Cortizo, G., Cabot, M. I., Cubero, M. T. G., Coca, M., Ferrari, M. D., & Lareo, C. (2020). Bioprocess intensification for isopropanol, butanol and ethanol (IBE) production by fermentation from sugarcane and sweet sorghum juices through a gas stripping-pervaporation recovery process. *Fuel*, *281*, 118593. https://doi.org/10.1016/j.fuel.2020.118593

Sajjad, Z., Gilani, M. A., Nizami, A.-S., Bilad, M. R., & Khan, A. L. (2019). Development of novel hydrophilic ionic liquid membranes for the recovery of biobutanol through pervaporation. *Journal of Environmental Management*, *251*, 109618. https://doi.org/10.1016/j.jenvman.2019.109618

Santoro, S., Galiano, F., Jansen, J. C., & Figoli, A. (2017). Strategy for scale-up of SBS pervaporation membranes for ethanol recovery from diluted aqueous solutions. *Separation and Purification Technology*, *176*, 252–261. https://doi.org/10.1016/j.seppur.2016.12.018

Santos, E. L. I., Rostro-Alanís, M., Parra-Saldívar, R., & Alvarez, A. J. (2018). A novel method for bioethanol production using immobilized yeast cells in calcium-alginate films and hybrid composite pervaporation membrane. *Bioresource Technology*, *247*, 165–173. https://doi.org/10.1016/j.biortech.2017.09.091

Setlhaku, M., Heitmann, S., Górak, A., & Wichmann, R. (2013). Investigation of gas stripping and pervaporation for improved feasibility of two-stage butanol production process. *Bioresource Technology*, *136*, 102–108. https://doi.org/10.1016/j.biortech.2013.02.046

Shin, C., Baer, Z. C., Chen, X. C., Ozcam, A. E., Clark, D. S., & Balsara, N. P. (2015). Block copolymer pervaporation membrane for in situ product removal during acetone–butanol–ethanol fermentation. *Journal of Membrane Science*, *484*, 57–63. https://doi.org/10.1016/j.memsci.2015.03.005

Si, Z., Cai, D., Li, S., Zhang, C., Qin, P., & Tan, T. (2019). Carbonized ZIF-8 incorporated mixed matrix membrane for stable ABE recovery from fermentation broth. *Journal of Membrane Science*, *579*, 309–317. https://doi.org/10.1016/j.memsci.2019.02.061

Silva, L. R., Guimarães, A. S., do Nascimento, J., do Santos Nascimento, I. J., da Silva, E. B., McKerrow, J. H., Cardoso, S. H., & da Silva-Júnior, E. F. (2021). Computer-aided design of 1,4-naphthoquinone-based inhibitors targeting cruzain and rhodesain cysteine proteases. *Bioorganic and Medicinal Chemistry*, *41*, 116213. https://doi.org/10.1016/j.bmc.2021.116213

Song, Y., Lee, Y. G., Cho, E. J., & Bae, H.-J. (2020). Production of xylose, xylulose, xylitol, and bioethanol from waste bamboo using hydrogen peroxicde-acetic acid pretreatment. *Fuel*, *278*, 118247. https://doi.org/10.1016/j.fuel.2020.118247

Song, Y., Nguyen, Q. A., Wi, S. G., Yang, J., & Bae, H.-J. (2017). Strategy for dual production of bioethanol and D-psicose as value-added products from cruciferous vegetable residue. *Bioresource Technology*, *223*(Part 1), 34–39. https://doi.org/10.1016/j.biortech.2016.10.021

Song, Y., Oh, C., & Bae, H.-J. (2017). Simultaneous production of bioethanol and value-added D-psicose from Jerusalem artichoke (*Helianthus tuberosus* L.) tubers. *Bioresource Technology*, *244*, 1068–1072. https://doi.org/10.1016/j.biortech.2017.08.079

Su, G., Wang, C., Tan, J. K., Zhang, C., & He, J. (2024). Enhanced butanol production from dough and okara waste through co-fermentation. *Renewable Energy*, *234*, 121157. https://doi.org/10.1016/j.renene.2024.121157

Trinh, L. T. P., Cho, E. J., Lee, Y. J., Bae, H.-J., & Lee, H.-J. (2013). Pervaporative separation of bioethanol produced from the fermentation of waste newspaper. *Journal of Industrial and Engineering Chemistry*, *19*(6), 1910–1915. https://doi.org/10.1016/j.jiec.2013.02.036

Trinh, L. T. P., Lee, Y.-J., Park, C. S., & Bae, H.-J. (2019). Aqueous acidified ionic liquid pretreatment for bioethanol production and concentration of produced ethanol by pervaporation. *Journal of Industrial and Engineering Chemistry*, *69*, 57–65. https://doi.org/10.1016/j.jiec.2018.09.008

Van Goethem, C., Naik, P. V., de Velde, M. V., Van Durme, J., Verplaetse, A., & Vankelecom, I. F. J. (2023). Stability of filled PDMS pervaporation membranes in bio-ethanol recovery from a real fermentation broth. *Membranes*, *13*(11), 863. https://doi.org/10.3390/membranes13110863

Van Hecke, W., & De Wever, H. (2017). High-flux POMS organophilic pervaporation for ABE recovery applied in fed-batch and continuous set-ups. *Journal of Membrane Science*, *540*, 321–332. https://doi.org/10.1016/j.memsci.2017.06.058

Van Hecke, W., Vandezande, P., Dubreuil, M., Uyttebroek, M., Beckers, H., & De Wever, H. (2016). Biobutanol production from C5/C6 carbohydrates integrated with pervaporation: experimental results and conceptual plant design. *Journal of Industrial Microbiology and Biotechnology*, *43*(1), 25–36. https://doi.org/10.1007/s10295-015-1717-3

Wagner, E. R., & Gasch, A. P. (2023). Advances in *S. cerevisiae* engineering for xylose fermentation and biofuel production: Balancing growth, metabolism, and defense. *Journal of Fungi*, *9*(8), 786. https://doi.org/10.3390/jof9080786

Wang, F., Shen, M., Li, Z., Zhu, Y., Li, J., Wang, N., Meng, H., & An, Q.-F. (2024). Slot-die coating of vinyl-PDMS composite membranes for alcohol perm-selective pervaporation. *Journal of Membrane Science*, *709*, 123103. https://doi.org/10.1016/j.memsci.2024.123103

Wang, Z., Xu, W., Gao, Y., Zha, M., Zhang, D., Peng, X., Zhang, H., Wang, C., Xu, C., Zhou, T., Liu, D., Niu, H., Liu, Q., Chen, Y., Zhu, C., Guo, T., & Ying, H. (2023). Engineering *Saccharomyces cerevisiae* for improved biofilm formation and ethanol production in continuous fermentation. *Biotechnology for Biofuels and Bioproducts*, *16*, 119. https://doi.org/10.1186/s13068-023-02356-6

Xin, F., Wang, C., Dong, W., Zhang, W., Wu, H., Ma, J., & Jiang, M. (2016). Comprehensive investigations of biobutanol production by a non-acetone and 1,3-propanediol generating *Clostridium* strain from glycerol and polysaccharides. *Biotechnology for Biofuels*, *9*, 220. https://doi.org/10.1186/s13068-016-0641-8

Xue, C., Liu, F., Xu, M., Zhao, J., Chen, L., Ren, J., Bai, F., & Yang, S.-T. (2016). A novel *in situ* gas stripping-pervaporation process integrated with acetone-butanol-ethanol fermentation for hyper n-butanol production. *Biotechnology and Bioengineering*, *113*(1), 120–129. https://doi.org/10.1002/bit.25666

Xue, C., Wang, Z.-X., Du, G.-Q., Fan, L.-H., Mu, Y., Ren, J.-G., & Bai, F.-W. (2016). Integration of ethanol removal using carbon nanotube (CNT)-mixed membrane and ethanol fermentation by self-flocculating yeast for antifouling ethanol recovery. *Process Biochemistry*, *51*(9), 1140–1146. https://doi.org/10.1016/j.procbio.2016.05.030

Xue, C., Yang, D., Du, G., Chen, L., Ren, J., & Bai, F. (2015). Evaluation of hydrophobic micro-zeolite-mixed matrix membrane and integrated with acetone–butanol–ethanol fermentation for enhanced butanol production. *Biotechnology for Biofuels*, *8*, 105. https://doi.org/10.1186/s13068-015-0288-x

Yang, Y., Sun, H., He, Y., Li, J., Wang, N., & An, Q.-F. (2024). Hydroxy salt induced CoAl-LDH nanosheet composite membrane for pervaporation desalination. *Desalination*, *592*, 118144. https://doi.org/10.1016/j.desal.2024.118144

Zhang, C., Pang, S., Lv, M., Wang, X., Su, C., Gao, W., Chen, H., Cai, D., Qin, P., & Tan, T. (2019). Reduction of wastewater discharge toward second-generation acetone–butanol–ethanol production: Broth recycling by the fermentation–pervaporation hybrid process. *Energy and Fuels*, *33*(2), 1210–1218. https://doi.org/10.1021/acs.energyfuels.8b03526

Zhang, C., Si, Z., Chen, B., Chen, C., Chen, H., Ren, W., Cheng, S., Li, S., Cai, D., & Qin, P. (2021). Co-generation of acetone-butanol-ethanol and lipids by a sequential fermentation using *Clostridia*

acetobutylicum and *Rhodotorula glutinis*, spaced-out by an *ex-situ* pervaporation step. *Journal of Cleaner Production*, *285*, 124902. https://doi.org/10.1016/j.jclepro.2020.124902

Zhang, H., Zhang, P., Wu, T., & Ruan, H. (2023). Bioethanol production based on *Saccharomyces cerevisiae*: Opportunities and challenges. *Fermentation*, *9*(8), 709. https://doi.org/10.3390/fermentation9080709

Zhang, Z., Xu, S., Wu, Y., Shi, S., & Xiao, G. (2021). Recent advances of pervaporation separation in DMF/H_2O solutions: A review. *Membranes*, *11*(6), 455. https://doi.org/10.3390/membranes11060455

Zhu, H., Li, X., Pan, Y., Liu, G., Wu, H., Jiang, M., & Jin, W. (2020). Fluorinated PDMS membrane with anti-biofouling property for *in-situ* biobutanol recovery from fermentation-pervaporation coupled process. *Journal of Membrane Science*, *609*, 118225. https://doi.org/10.1016/j.memsci.2020.118225

Zhu, H., Liu, G., Yuan, J., Chen, T., Xin, F., Jiang, M., Fan, Y., & Jin, W. (2019). *In-situ* recovery of bio-butanol from glycerol fermentation using PDMS/ceramic composite membrane. *Separation and Purification Technology*, *229*, 115811. https://doi.org/10.1016/j.seppur.2019.115811

Pertanika J. Sci. & Technol. 33 (6): 2589 - 2613 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Optimisation and Predictive Modelling of Natural Frequencies on Carbon/Glass Hybrid Composite Laminates

Abdul Azim Taredi, Siti Mariam Abdul Rahman, Mohd Nor Azmi Ab Patar and Jamaluddin Mahmud*

School of Mechanical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

ABSTRACT

A hybrid laminated composite is a laminate formed by integrating composite layers from various fibre types to achieve optimal properties. Nonetheless, considerable knowledge remains to be acquired regarding the vibration behaviour associated with the hybridisation of composite laminates. In structural design, natural frequency is a critical factor for preventing resonance, which may result in significant structural failure. This study aims to assess the inherent natural frequency response of hybrid composite laminates under free vibration, influenced by varying plate thicknesses, layer fractions, and orientation angles. Finite element models were developed using a commercial finite element software, ANSYS, to precisely characterise the natural frequencies of hybrid composite laminates under free vibration. The design of experiments was employed to identify 17 case study runs and to assess significant factors, with a comprehensive examination of each factor's impact on natural frequencies conducted through modal analysis. Given the limited dataset, employing techniques such as cross-validation with response surface methodology (RSM) and artificial neural networks (ANN) enhances the reliability of performance assessment for the model. Optimisation was conducted utilising RSM via analysis of variance, while ANN serves as a tool to ascertain data accuracy. The accuracy and robustness of the models are corroborated by a comparison of predictions from finite element analysis and RSM, demonstrating a strong correlation with percentage errors of 16 and 10% for ANN, respectively.

Keywords: ANSYS, artificial neural network, finite element analysis, hybrid composite laminate, natural frequency, response surface methodology

ARTICLE INFO

Article history:
Received: 23 November 2024
Accepted: 18 May 2025
Published: 08 October 2025

DOI: https://doi.org/10.47836/pjst.33.6.09

E-mail addresses:
2024631962@student.uitm.edu.my (Abdul Azim Taredi)
mariam4528@uitm.edu.my (Siti Mariam Abdul Rahman)
azmipatar@uitm.edu.my (Mohd Nor Azmi Ab Patar)
jm@uitm.edu.my (Jamaluddin Mahmud)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

Plate and shell structures are extremely significant and efficient load-bearing entities in various mechanical, aerospace, and civil engineering applications. Being thin-walled structures, they enable the designer to achieve substantial weight savings. The past few decades have witnessed the replacement of traditional materials with multi-layered composite structures such as laminates and sandwich panels, primarily due to their high stiffness-to-weight and flexural rigidity-to-weight ratio, respectively (Sai Vivek, 2016). Vibration analysis of rectangular plates has received considerable attention and has been the subject of numerous studies, which are well documented by Leissa (1973), who classified the accurate results for free vibration frequencies of all 21 combinations of classical boundary conditions for rectangular plates. Free vibration of isotropic and orthotropic rectangular plates with mixed boundary conditions is investigated by Mirziyod et al. (2019). Conducting physical experiments can be expensive, particularly when working with large-scale systems or processes. Simulation allows for the modelling of these systems without the need for physical experiments, which can significantly reduce the cost effectively. This is relevant to Industry 4.0, where a global survey with more than 2,000 respondents suggests that approximately 5% of a company's annual revenue will be invested in digitalisation projects (de Paula Ferreira et al., 2020). Over the past few decades, finite element modelling (FEM) has been used more often to support engineering projects because of its inexpensive prices, reduced project elaboration time, and improved computational and data-processing capabilities (Müzel et al., 2020). Solid elements are the least employed in composites since they necessitate a multi-layered model or an expensive, time-consuming full-size construction, which makes them impractical (Chisena et al., 2021).

In recent years, the integration of composite materials has gained significant attention across various industries, particularly due to their superior structural performance and vibration characteristics. Hybrid composites, which are combinations of various fibre materials, have exhibited significant potential in aerospace, automotive, construction and consumer sectors. The natural frequency refers to the dynamic properties, exhibiting the level of vibration in relation to the mass and stiffness of a structure (Alam et al., 2020; Norman & Mahmud, 2021). Comprehending natural frequencies and mode shapes is essential for averting sudden and catastrophic failures, as each structure possesses distinct resonant frequencies (An et al., 2019; Ratnaparkhi & Sarnobat, 2013). This paper considers the frequencies of hybrid composites of rectangular full plates subjected to free vibration. During vibration, spots on the plate undergo small displacements in the direction perpendicular to the plane of the plate (Warburton, 1954).

Composite plates play a vital role, particularly in hybrid laminates, where the stacking sequence, fibre orientation, and material selection profoundly influence dynamic response (B. S. Reddy et al., 2013; Jadee et al., 2020). Recent studies have delved into various

factors influencing the vibration behaviour of hybrid composites. For instance, Li et al. (2024) performed an extensive analysis of carbon/glass hybrid composites laminated plates under various boundary conditions, emphasising the influence of hybrid ratio, stacking sequences, and fibre orientations on their vibrational properties. Ma and Gao (2024) investigated the free vibration of hybrid laminated thin-walled cylindrical shells composed of multilayer functionally graded carbon nanotube-reinforced composite plies, highlighting the significance of material gradation in dynamic performance. A study by Sreekanth et al. (2023) discussed the initiation of delamination that had caused the major failure of laminated composite structures, affected by the dynamic characteristics and vibration properties. Likewise, Bulut et al. (2016) and Quanjin et al. (2022) conducted physical tests and numerical simulations for free vibrations of Kevlar/glass epoxy laminated composite plates. Arvinda Pandian and Jailani (2021) investigated the mechanical and thermal properties of hybrid epoxy laminates reinforced with jute and linen fabrics along with fumed silica filler, highlighting how material composition influences composite behaviour. Erkli‡ et al. (2015) also examined the influence of hybridisation on natural frequency through experiment and simulation. Norman et al. (2018) investigated the free vibrations of the laminated composite beam, analysing the influence of different lamination schemes using ANSYS software. Shi et al. (2020) developed an analytical approach for examining free vibrations of cross-ply Glass/epoxy composite laminates with shallow shells by incorporating a semi-analytic method into arbitrary classical and elastic boundary conditions. There are also studies related to free vibrations of hybrid laminated composite plates. A study showed that the matrix material, hybridisation, and laminate stacking sequence affect the natural frequencies and mode shapes. These studies cover diverse areas, including the exploration of the impact of varying fibre volume fraction ratios, as demonstrated by Pushparaj and Suresha (2016). Finite element analysis (FEA) was used by Crawley (1979) in the numerical approach, and the outcomes of the experiments were compared. However, there is still much to learn about natural frequencies in hybrid composites, even though previous studies have illuminated a number of these topics (Safri et al., 2018).

Therefore, this paper aims to investigate the influence factors of natural frequency response to hybrid composite, and the optimum factors are also determined by implementing the RSM. Optimising process factors is a significant challenge in experimental work, crucial for conserving resources, reducing costs, and achieving improved outcomes. To do so, the advanced statistical tools must be adopted within the investigation. RSM is a sophisticated design of experiments (DOE) technique that involves statistical modelling to develop and analyse processes, with the goal of optimising the desired response influenced by multiple input factors (Alam et al., 2020; Bezerra et al., 2008; Shokuhfar et al., 2008; Swamy et al., 2014). Using the DOE approach reduced the number of necessary experiments while still

enabling the collection of significant amounts of data. When compared to the sequential variation of one factor at a time, this approach enabled the concurrent exploration of multiple factors, resulting in a significant reduction in the time and resources required (Antony, 2014). Implement the Box-Behnken design (BBD) technique; the experimental layout was created with three factors, each with three levels. The BBD was proposed by George E. P. Box and Donald W. Behnken in 1960. Their objective was to develop an effective experimental design methodology capable of accommodating second-order (quadratic) models without necessitating the extensive number of experiments required for a complete three-level factorial design (Box & Behnken, 1960). Thus, the design is a great option for maximising the natural frequency of hybrid composite laminates because it works particularly well for building second-order polynomial models and investigating quadratic response surfaces (Triefenbach, 2008). Thus, BBD provides the most information with their greater symmetry and rotatability and requires fewer experimental runs (Antony, 2014). This design offers a thorough understanding of the interactions between variables by simultaneously varying multiple factors, which produces more robust and trustworthy results. Nonetheless, although RSM provides interpretable models, its prediction accuracy may be constrained when addressing highly nonlinear and intricate connections frequently encountered in hybrid composite structures (Saravanakumar et al., 2024).

The last few decades have focused a lot of emphasis on the integration of composite materials in a variety of applications, especially regarding their structural performance and vibration characteristics. According to a recent study, the use of ANN modelling is becoming more common and may even reduce process complexity (Subrahmanyam et al., 2021). For instance, studies have investigated using ANN as a failure structure prediction tool (Kumar C & Swamy, 2021; Norman et al., 2022; Sreekanth et al., 2023). Neural networks are constructed by connecting neurons through weighted interconnections, and the most critical step is determining these weights, a process known as training. In developing the ANN model, the Levenberg-Marquardt (LM) algorithm has gained popularity among studies, as it has proven efficiency in training feed-forward networks for vibration analysis of composite structures. Previous studies have demonstrated the algorithm's effectiveness in predicting vibration response under varying environmental conditions with high accuracy (Kallannavar et al., 2020). Other studies developed an ANN model using a multilayer perceptron (MLP) with backpropagation to predict the natural frequencies of laminated composite plates with higher accuracy (M. R. S. Reddy et al., 2012). However, smaller datasets may pose a risk of overfitting if not managed properly. In this context, Arunachalam et al. (2024) conducted experiments and simulation analysis on predicting the hardness and strength of the polymer composite by applying silane treatment. Optimisation of RSM was done by indicating the influence factor and ANN to predict the flexural strength and hardness. Even though the small dataset produced a high of 95% accuracy of predictions by using an

ANN. Mallick et al. (2024) investigate an electroless nickel-phosphorus (Ni–P) coating's ability to withstand corrosion on a copper substrate using BBD and ANN. It has 17 data sets given by BBD, where 70% of the data have been used for input, 15% for testing and 15% for validation. BBD produces 87% accuracy, and 88% for ANN.

In this study, the plate thickness, layer fraction, and lamination schemes have been considered as the three factors to be optimised using BBD to achieve optimal natural frequencies of a hybrid composite laminate. The natural frequencies provide insight into the vibrational behaviour and structural integrity of the carbon/glass under various conditions. Additionally, a predictive model can be useful to predict the natural frequencies of this hybrid composite across multiple values of input factors. To identify the significant factors in this process, analysis of variance (ANOVA), a robust statistical tool, has been employed to analyse residuals, detect outliers, and evaluate the validity of the developed model. Furthermore, a more precise predictive model for identifying the best-fit natural frequency responses of the hybrid composite laminates has been established using an ANN.

METHODOLOGY

Mesh Convergence Analysis and Numerical Validation

Mesh Convergence Analysis

In finite element analysis, mesh convergence analysis is essential for determining the optimal mesh size to achieve a balance between accuracy and computational efficiency (Samsudin & Mahmud, 2015). To verify that the simulation results converge to a stable solution, it assesses the sensitivity of numerical outcomes to variations in mesh size. As a result, a range of mesh sizes was employed. 1×2, 2×4, 3×6, 4×8, 5×10, 6×12, 7×14, 8×16, and 16×32. The mesh employs a quadrilateral configuration on a model measuring 150 mm in length and 75 mm in width, utilising the shell element type 8 node 281. This stage encompassed a range of angles from 0° to 90° and employed symmetric angle lamination [θ°/-θ°/θ°/-θ°]s as the stacking configuration. This aims to ascertain the behaviour of natural frequencies and the deformed shape.

Numerical Validation

The purpose of numerical validation is to confirm the accuracy of free vibration analysis for a hybrid laminated composite by validating the finite element model using ANSYS *Parametric Design Language (*APDL) (v16.0, SAS IP, Inc., USA). The validation procedure is founded on the work of (Pushparaj & Suresha, 2016) and the study conducted by Norman and Mahmud (2021). The hybrid composite consists of E-glass/T300 carbon epoxy plates (Pushparaj & Suresha, 2016) measuring 150 mm in length and 75 mm in width. Figure 1 illustrates the dimensions and clamped-free-free-free (CFFF) boundary condition of the

plate. The ply thickness of the T300 carbon fibre lamina is 0.13 mm, whereas the E-glass fibre lamina has a ply thickness of 0.15 mm. The plate consists of 8 layers, which were divided into 400 elements using eight-node quadrilateral shell elements, namely SHELL 281, and the material properties are tabulated in Table 1. The lamination angle is [0g/+45c/-45c/90g]s and [0c/+45g/-45g/90c]s was examined. Thus, the first five mode shapes and natural frequencies of the plate are used in the Block Lanczos mode extraction method.

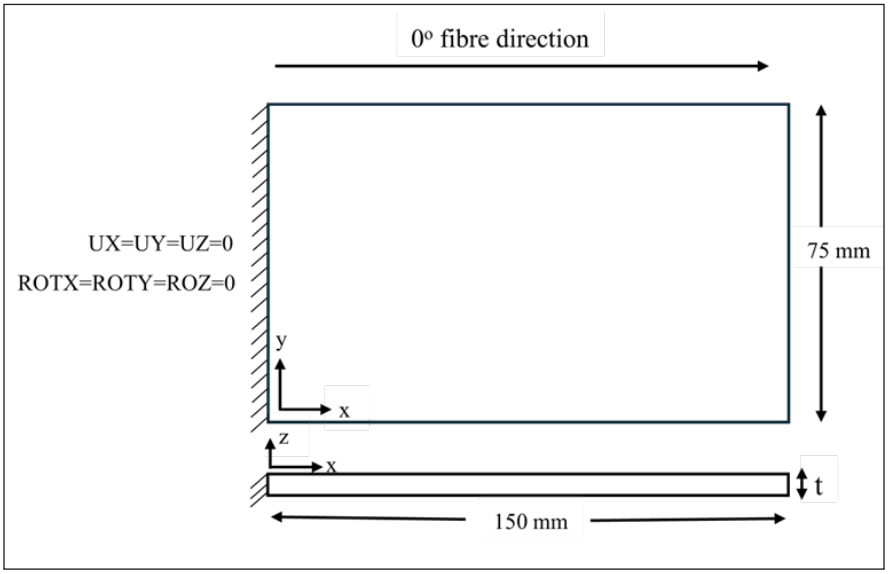


Figure 1. Rectangular plate configuration
Note. UX = Displacement in x-direction; UY = Displacement in y-direction; UZ = Displacement in z-direction; ROTX = Rotation about x-axis; ROTY = Rotation about y-axis; ROTZ = Rotation about z-axis; t = Thickness of the laminate, which is varied depending on the number of laminas

Table 1
Material properties of T-300 carbon epoxy (Samsudin & Mahmud, 2015) and E-glass epoxy (Pushparaj & Suresha, 2016)

T-300 carbon epoxy	Values	E-glass epoxy	Values
E_1	132.4 GPa	E_1	44.93 GPa
$E_2 = E_3$	10.76 GPa	$E_2 = E_3$	14.04 Gpa
$v_{12} = v_{13}$	0.24	$v_{12} = v_{13}$	0.2481
v_{23}	0.49	v_{23}	0.3775
$G_{12} = G_{13}$	5.645 GPa	$G_{12} = G_{13}$	5.263 Gpa
G_{23}	3.38 GPa	G_{23}	5.101 Gpa
ρ	1,600 kg/m^3	ρ	2,081 kg/m^3

Note. E_1 = Longitudinal Young's modulus; E_2 = Transverse Young's modulus; E_3 = Through-thickness Young's modulus; v_{12} = Poison ratio in 1-2 plane, v_{13}= Poison ratio in 1-3 plane; v_{23} = Poison ratio in 2-3 plane; G_{12} = Shear modulus in 1-2 plane; G_{13} = Shear modulus in 1-3 plane; G_{23} = Shear modulus in 2-3 plane; ρ = Density

Set Up Factors Using the DOE and Vibration Analysis

Statistical optimisation was conducted using the BBD implemented in Design Expert (22.0.6 64-bit) to optimise factors for symmetric configurations as shown in Table 2. At least three continuous elements are needed for this cubic design, which is defined by the midpoints of the edges of a multi-dimensional cube. In this study, the effects of three numerical factors, namely fibre angle orientation (θ°), layer fraction (Lf, %) and plate thickness (t, mm), on the natural frequencies of the hybrid composite laminate were investigated. The laminate consists of 8 layers arranged in a symmetric lamination scheme of [θ°/-θ°/θ°/-θ°]s. For variation, the minimum and maximum levels were set as follows: θ = 0° to 45°, Lf = 0 (0%) to 1 (100%) and t = 0.5 to 3 mm, corresponding to fibre angle orientation, layer fraction and plate thickness, respectively. The layer fraction represents the percentage of carbon/epoxy lamina: a value of '0' indicates that all eight layers are made of glass/epoxy, '0.5' indicates that there are four layers of carbon/epoxy and four layers of glass/epoxy, and '1' indicates all eight layers are carbon/epoxy. In all cases, the stacking sequence follows the [θ°/-θ°/θ°/-θ°]s configuration.

The finite element model is constructed as illustrated in Figure 1 using APDL (v16.0, SAS IP, Inc., USA) for the numerical validation. The properties of the materials, as well as additional DOE-set factors, are applied. The shell 281 elements are the same as in the first stage, and CFFF for boundary conditions are used.

Table 2
Box Behnken configuration

Variables	Low (-)	Middle (0)	High (+)
A: Angle orientation (θ°)	0	22.50	45
B: Layer fraction (%)	0	0.50	1
C: Plate thickness (mm)	0.50	1.75	3

Optimisation Using RSM

In this phase, the link between controllable input elements and the generated response surfaces is characterised using RSM. The methodical examination of the impacts of several variables and their interactions, RSM, which is backed by Design Expert software (22.0.6 64-bit), allows for a thorough study of input-output data and makes it easier to optimise the response variable. ANOVA is used in the BBD model to evaluate each component's statistical significance and how it interacts with the response. ANOVA is crucial for determining which factors have a significant influence on the response and for confirming the model's robustness. Regression equations and contour plots are created to show and analyse the relationships between input and output variables based on findings from previous stages.

Prediction of Natural Frequency Using ANN

The first step in using an ANN to predict natural frequencies was gathering and preprocessing response data from simulation samples. The natural frequencies of DOE samples were examined by using ANSYS. This ensures the neural network model can operate with the data, which encompasses a variety of factors and their corresponding responses and has been normalised by Malik and Arif (2013). This study's ANN model was trained via the LM algorithm, a powerful optimisation method recognised for its swift convergence and low mean squared error (MSE) in small to medium datasets (Kallannavar et al., 2020). The dataset produced by the BBD was randomly partitioned into 70% for training, 15% for validation, and testing to ensure model generalisation and mitigate overfitting (Boukarma et al., 2023; Subrahmanyam et al., 2021). The model's prediction accuracy was evaluated using Equation 1, MSE and Equation 2, root mean squared error (RMSE) metrics, where MSE measures the average of the squares of the differences between the predicted $\hat{y}_i$ and actual y_i values. Thus, RMSE provides an error value in the same units as the original outputs, measuring how accurately the prediction model performs. The input values were the predetermined variables for BBD, which are as follows: angle orientation (A), layer fraction (B), and plate thickness (C), as shown in Figure 2.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \quad [1]$$

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

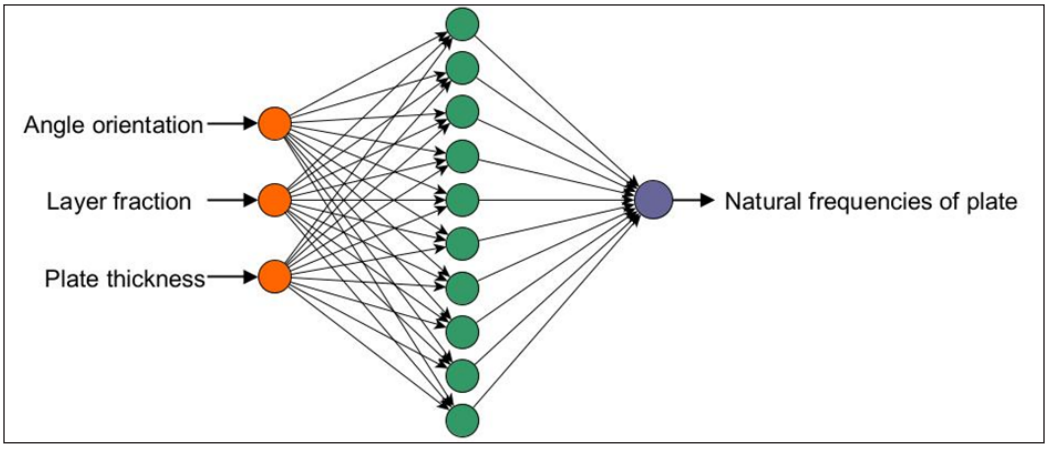


Figure 2. Architecture of the proposed artificial neural network predictive model

RESULTS AND DISCUSSION

Mesh Convergence and Numerical Validation

Convergence Analysis

The results obtained from the mesh convergence analysis are shown in Figure 3. The different mesh sizes in hybrid laminated composites are responsible for the slight differences in natural frequencies. The results plateau at mesh sizes larger than 8x16, or 128 elements, suggesting that the effect on natural frequencies is negligible. Therefore, for the ensuing analyses, an 8 × 16 mesh was used for this study. Interestingly, it is worth noting that these findings are particular to dataset mode number 5.

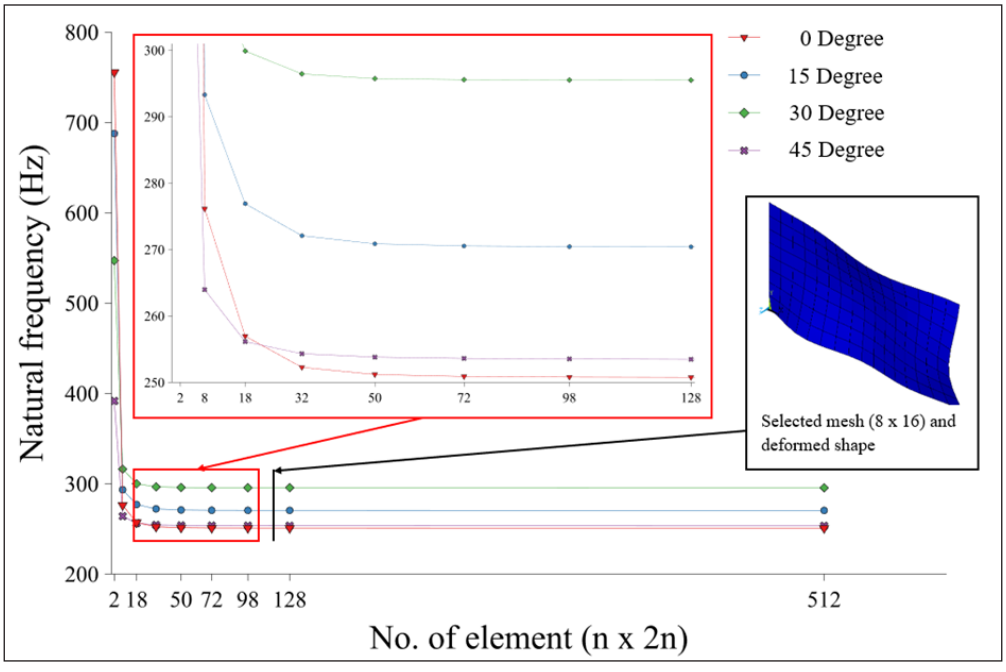


Figure 3. Convergence analysis of the mesh for the hybrid composite plate [0g/+45c/-45c/90g]s

Numerical Validation

Table 3 compares the results of this simulation to (Pushparaj & Suresha, 2016). The acquired natural frequencies and the mode shapes of the hybrid T300 carbon/E-glass epoxy were compared. The data indicates a modulus value with the maximum true error recorded at under 5.02%. The findings indicate that changes in lamina sequences result in differences in mode shapes, although they resemble those identified in the previous study. The literature indicates that this study demonstrated a significant level of agreement in the similarity response.

Table 3
Comparison of simulated natural frequencies and mode shapes with literature

Lamination scheme	Mode	Pushparaj and Suresha (2016)	Present study	True error (%)
		Numerical simulation (Hz)	Numerical simulation (Hz)	
[0g/45c/-45c/90g]s	1 Bending	37.374	37.465	0.24
	1 Torsion	165.05	162.74	1.40
	2 Bending	231.26	232.18	0.40
	2 Torsion	550.26	522.61	5.02
	1 Combined	645.65	648.03	0.37
[0c/45g/-45g/90c]s	1 Bending	54.697	53.277	2.60
	1 Torsion	137.08	139.62	1.85
	2 Bending	341.24	332.42	2.58
	2 Torsion	495.24	497.9	0.54
	1 Combined	654.26	670.86	2.54

DOE Factors and Vibration Analysis

This result was generated using the BBD, which was employed to classify the three factors of the experiment. It set the case study that involved developing 17 simulations, all of which had to be completed using FEA. This result contains angle orientation [θ°/-θ°/θ°/-θ°] s for symmetric with fixed variable layer fraction 0,0.5, and 1, then thickness from 0.5 to 3 mm, and angle of orientation 0° until 45°. For a hybrid arrangement, the layup is [C, G, C, G, G, C, G, C], where C is carbon and G is glass. Table 4 shows the completed value of FEA for the symmetric.

Table 4
Natural frequencies for symmetric laminate

Run	Factor 1 A: Angle of orientation (Degree)	Factor 2 B: Layer fraction (%)	Factor 3 C: Plate thickness (mm)	FEA	RSM	ANN	Error of RSM (%)	Error of ANN (%)
1	0	0.5	0.50	27.18	24.31	30.11	10.56	10.78
2	0	0.5	3.00	162.77	161.75	162.9	0.63	0.08
3	22.5	0	3.00	88.06	86.59	88.89	1.67	0.94
4	45	0	1.75	38.00	38.00	37.97	0.00	0.09
5	0	1	1.75	114.38	122.38	114.4	7.00	0.02
6	45	0.5	0.50	12.95	15.07	13.94	16.33	7.63
7	22.5	0.5	1.75	72.50	70.50	72.70	2.75	0.14
8	22.5	0.5	1.75	72.50	70.50	72.70	2.75	0.14
9	22.5	0.5	1.75	72.50	70.50	72.70	2.75	0.14
10	22.5	0.5	1.75	72.50	70.50	72.70	2.75	0.14
11	22.5	0	0.50	14.71	15.09	15.00	2.55	1.95

Table 4 *(continue)*

Run	Factor 1 A: Angle of orientation (Degree)	Factor 2 B: Layer fraction (%)	Factor 3 C: Plate thickness (mm)	FEA	RSM	ANN	Error of RSM (%)	Error of ANN (%)
12	22.5	1	3.00	160.54	156.05	161.00	2.80	0.29
13	45	0.5	3.00	76.93	80.89	83.16	5.15	8.10
14	45	1	1.75	54.94	57.96	55.08	5.49	0.26
15	0	0	1.75	58.69	63.67	54.95	8.50	6.37
16	22.5	1	0.50	26.93	24.29	29.77	9.81	10.53
17	22.5	0.5	1.75	72.50	70.50	72.60	2.75	0.14

Note. FEA = Finite element analysis; RSM = Response surface methodology; ANN = Artificial neural network

The modal analysis employing the Block Lanczos method yielded multiple mode shapes and natural frequencies, derived from a specified number of modes for extraction. Each mode shape displayed distinct deformation types, such as torsion, bending, or a combination of both. The selection of the first natural frequency is based on the mode shape of the first bending mode, as shown in Figure 4 (a). The second natural frequency aligns with the

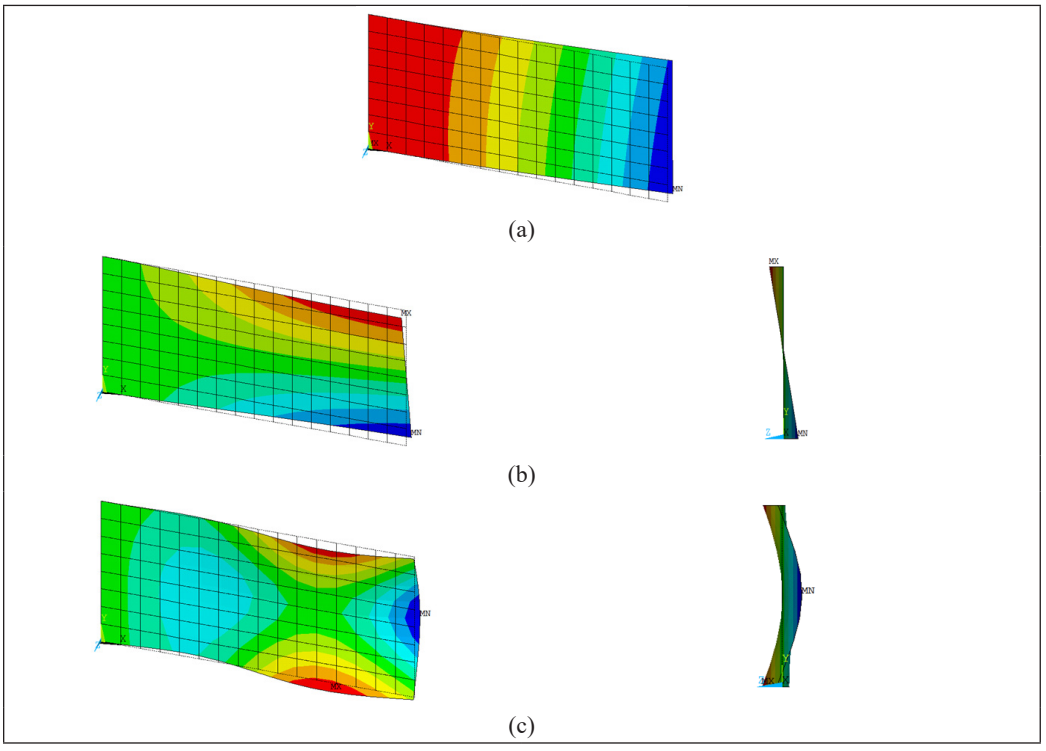


Figure 4. Classification of mode: (a) 1 bending, (b) 1 torsion, and (c) 1 combined
Note. X = x-axis; Y = y-axis; Z = z-axis; MN = Minimum; MX = Maximum

initial torsion mode depicted in Figure 4 (b), and the third natural frequency represents the combined torsion and bending mode, illustrated in Figure 4 (c). This deformation is verified by Pushparaj and Suresha (2016) and is comparable to the work of Merzuki et al. (2019), Njim et al. (2021), Norman et al. (2018), and Norman and Mahmud (2021). The mode shapes were deformed in exact alignment in accordance with both the experimental and numerical findings, demonstrating a remarkable level of concordance in this study.

Optimisation Using RSM

ANOVA based on the BBD was performed with Design Expert software to check the adequacy and significance of the model coefficients (Mallick et al., 2024; Mengistu et al., 2024; Rosaidi et al., 2023). An actual Equation3 was formulated through the ANOVA after the evaluation of the FEA model. This equation yielded the prediction of natural frequencies. Thus, Equation 3 is for a symmetric case (f_1) where A is the angle of orientation, B is the layer fraction, and C is the plate thickness. Table 5 shows the efficacy of the RSM model in predicting natural frequencies, with most of the percentage error comparatively minimal. The maximum error in the symmetric laminate does not surpass 17%, indicating a high level of accuracy at samples with an average percentage error of 4.96%. The RSM model performs well at two-factor interaction (2FI), especially regarding smaller percentage errors, thereby validating its use in vibration analysis of composite structures. High coefficients of determination, $R^2 = 1$, indicate a good correlation between the measured and predicted response when it is closely aligned (Boukarma et al., 2023; Polat & Sayan, 2019). If the *p*-value of any process factor or its interactions is below 0.05, it can be concluded that those factors or their interactions are significant (Mallick et al., 2024; Mengistu et al., 2024; Rosaidi et al., 2023).

$$f_1(symmetric) = -11.44205 + 0.543536A + 16.52975B + 42922.49C - 0.861311AB - 636.54133AC + 24104BC \quad [3]$$

where f_1 = a symmetric case, A = angle of orientation, B = layer fraction, and C = plate thickness.

A substantial *F*-value and minimal *p*-value indicate that each term exerts a significant influence (Mat Jusoh et al., 2024; Polat & Sayan, 2019). The model's *p*-value of under 0.0001 and *F*-value of 286.35 for the symmetric laminate in Table 5 indicate a substantial level of statistical significance. Thus, the lack of fit values was found to be insignificant. The high R^2 value of 0.9942, signifying that 99.42% of the variability in the natural frequencies is

Table 5
Result of the two-factor interaction model for a symmetric laminate

Source	Sum of squares	Df	Mean square	*F*-value	*p*-value	
Model	30375.74	6	5062.62	286.35	<0.0001	Significant
A - Angle orientation	4058.65	1	4058.65	229.56	<0.0001	
B - Layer fraction	3094.05	1	3094.05	175.00	<0.0001	
C - Plate thickness	20657.63	1	20657.63	1168.43	<0.0001	
AB	375.57	1	375.57	0.00	0.0010	
AC	1282.03	1	1282.03	21.24	<0.0001	
BC	907.82	1	907.82	72.51	<0.0001	
Residual	176.80	10	17.68	51.35	<0.0001	
Lack of fit	176.80	6	29.47			
Pure error	0.00	4	0.0000			
Cor total	30552.54	16				

Note. R^2 = 0.9942, R^2(Adj) = 0.9907, R^2(Pred) = 0.9729, Adeq precision = 54.3630

accounted for by the model, enhances its credibility. The model's robustness and predictive capability are validated by the adjusted R^2 (0.9907) and predicted R^2 (0.9729) values. The adequate precision of 54.3630 signifies a high signal-to-noise ratio, which is beneficial for model accuracy. With an exceptionally high *F*-value of 1168.43, plate thickness (C) stands out as the most influential factor, indicating its predominant impact on natural frequencies. The angle orientation (A) and layer fraction (B) are highly significant, with *F*-values of 229.56 and 175.00, respectively, and they substantially contribute to the variation in natural frequencies. The interaction terms (AB, AC, and BC) are significantly notable; the interaction (*F*-value = 72.51) between plate thickness and layer fraction (AC) highlights the combined effect of these two factors on the natural frequencies.

Figure 5 displays the diagnostic plots typically used to examine the performance of the quadratic model, and to evaluate the ANOVA model (Boukarma et al., 2023). The errors are concentrated along the straight line near the central position, which is predominantly normally distributed. Nevertheless, several residuals exceed the red control limits in Figures 5(b) and 5(d). These points signify regions where the model exhibits overfitting, but the overall randomness of the scatter indicates that the model's assumptions are not severely violated, and keeping the predictions for natural frequencies is generally reliable (Sathishkumar et al., 2022). Figure 5(c) presents a comparison between the actual response values and their corresponding predicted response values generated by the model. The graph of predicted values compared to actual values is significant as it allows the identification of a value or a set of values that the model does not predict accurately.

Figure 6 illustrates the impact of interactions on the response. To accurately identify this effect, the response surface was plotted as a function of two variables, while maintaining the others at their central values. Figures 6(a)-(c) clearly demonstrate that the natural

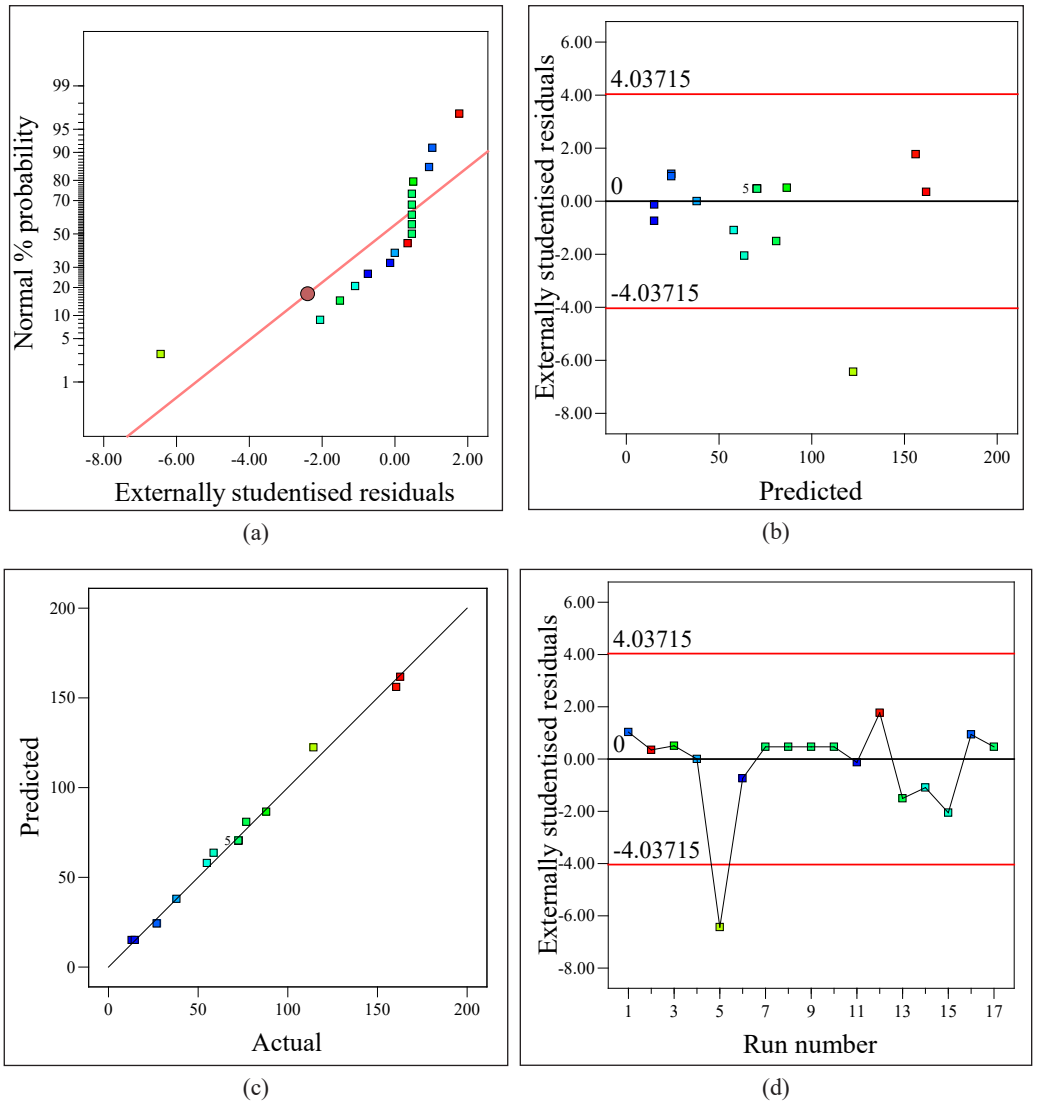
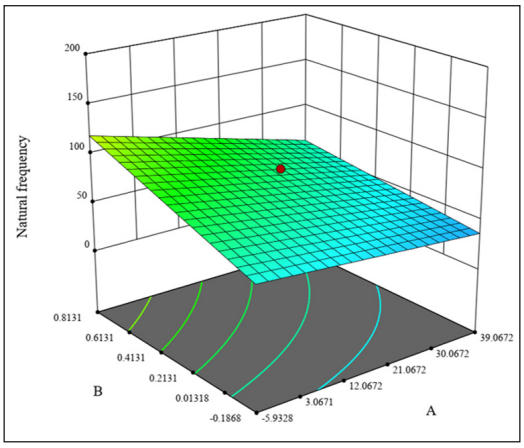
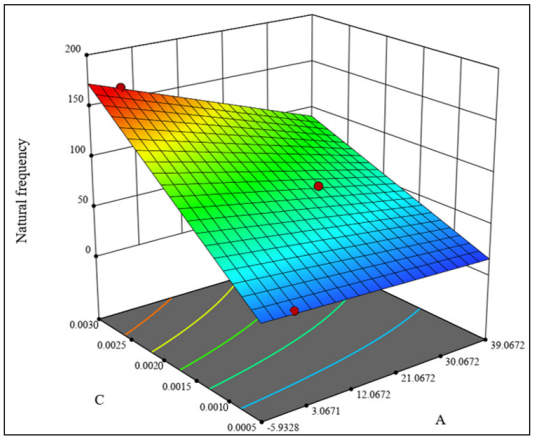


Figure 5. Diagnostic graph of (a) normal plot of residuals, (b) plot of residuals vs predicted, (c) regression analysis of predicted vs actual, and (d) residuals vs run number distribution

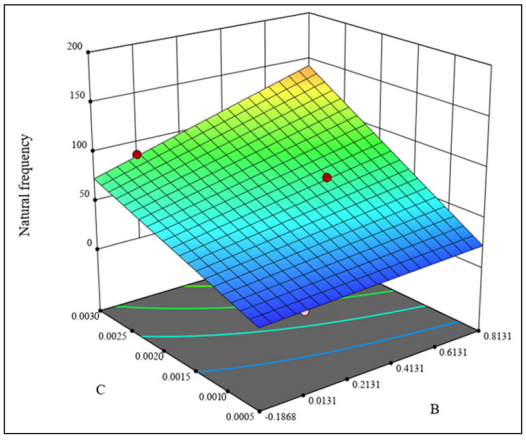
frequencies increase with the thickness of the plate. The interactions among the factors can be analysed from the figure, which illustrates that the second-order three-dimensional (3D) response surface plot, along with the contour plot of natural frequencies, corresponds to layer fraction with a lower angle orientation. Figure 6(c) provides a significant influence of the interaction between both factors on the natural frequencies, with colours transitioning from blue, indicating the low-response zone, to red, signifying optimal response. These results align with the findings of (Mat Jusoh et al., 2024). (Boukarma et al., 2023) proposed



(a)



(b)



(c)

Figure 6. Response surface plots representing the interaction of: (a) angle orientation (A) and layer fraction (B); (b) angle orientation (A) and plate thickness (C); and (c) layer fraction (B) and plate thickness (C)

the solution to ascertain which factors are more significant than others by selecting two points at the minimum and maximum levels on the graph. The first scenario features a minimum plate thickness and a maximum layer fraction (C = 0.0005m, B = 1%), while the second scenario is the inverse (C = 0.003m, B = 0%). The model predicts responses of 26.9 and 88.1 Hz for the two points, respectively.

Prediction of Natural Frequency Using ANN

The regression values are significant as they indicate the model's adequacy in fitting the data. A regression value approaching 1 signifies that the model exhibits high accuracy in predicting natural frequencies. The efficacy of ANN models for predicting natural frequencies in symmetric systems was summarised in Table 6. The ANN model attained a high R^2 (>0.99) and low RMSE (~2.8%), demonstrating superior prediction performance relative to traditional regression techniques. The efficacy of LM and data partitioning techniques has been continuously corroborated in the modelling of composite materials and the prediction of structural dynamics to prevent overfitting (Pratap et al., 2024). The hybrid composite laminate model exhibits a low training MSE of 0, a validation MSE of 35.3960, and a test MSE of 25.6818, indicating exceptional accuracy, supported by high R values of 1.0000 for training, 0.9971 for validation, and 0.9986 for testing.

Table 6
Model summary of an artificial neural network

Type	Mean squared error (MSE)	Regression (*R*)
Training	0.0000	1.0000
Validation	35.3960	0.9971
Test	25.6818	0.9986

Figure 7 illustrates the performance graphs of the ANN. The MSE evaluates the performance of the ANN over multiple epochs for training, validation, and test datasets. The optimal validation performance occurred at epoch 3, yielding a MSE of 35.3960. Effective learning is demonstrated by a steady decrease in training error, indicated by the blue line, as the number of epochs increases. This trend aligns with the findings of (Kumar C & Swamy, 2021), who predict the fatigue life of glass fibre, and the study by Mallick et al. (2024) though with a slightly different trajectory in this finding. Result shows the best validation performance for a total of five epochs with stagnant training and gradual increase in both validation and test MSE, like Arunachalam et al. (2024). The validation error on the green line, indicative of the model's generalisation to novel data, initially decreases before reaching a plateau. The test error on the red line, which evaluates the model's generalisation capability, exhibits a similar pattern.

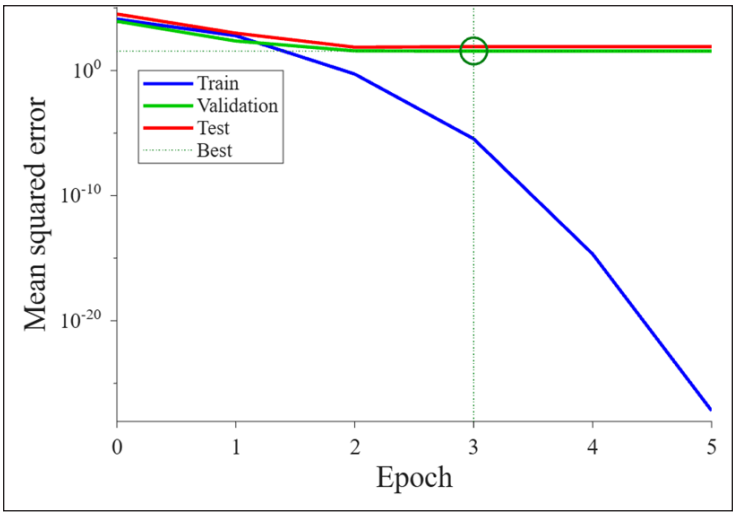


Figure 7. Performance plot of the artificial neural network
Note. The green circle highlights the best performance that happened at Epoch =3

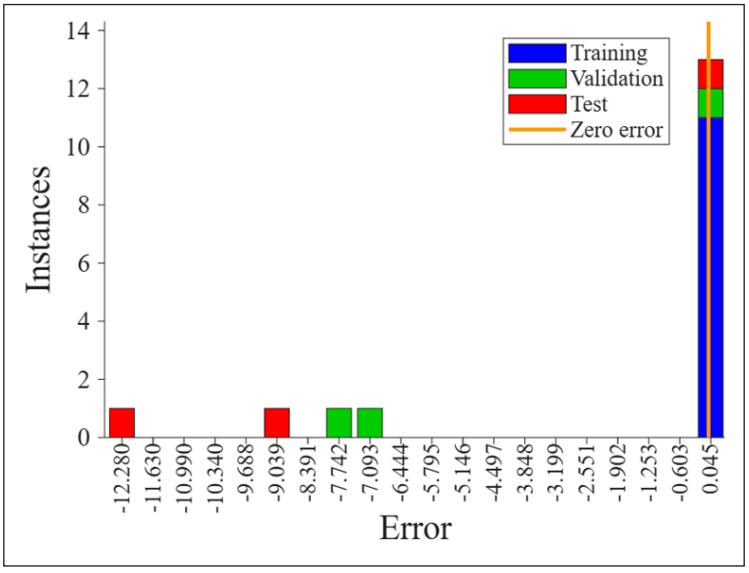


Figure 8. The value of the error histogram
Note. The vertical line in orange colour highlights the intersection with the x-axis at error = 0

The error histogram serves as a crucial diagnostic instrument for assessing the reliability and performance of predictive models. The distribution of errors across multiple datasets provides insights into the model's accuracy, generalisability, and potential areas for enhancement. Figure 8 presents error histograms with 20 bins corresponding to the training, validation, and testing phases of the network. The analysis indicates that the

network's predictions exhibit high precision and accuracy, with errors ranging from 0.03 to -0.5 within a narrow interval. Notable errors occur within the negative range, particularly between -12 and -6. The substantial negative errors suggest that, in some instances, the model significantly underestimates the target values. The distribution of test and validation errors across the bins indicates potential effective generalisation of the model; however, further analysis is required to assess overfitting.

Figure 9 presents regression analysis plots for the ANN models employed in predicting natural frequencies. The training dataset exhibits a correlation coefficient of 1, indicating robust model performance throughout the training phase. The test dataset attains an *R* value

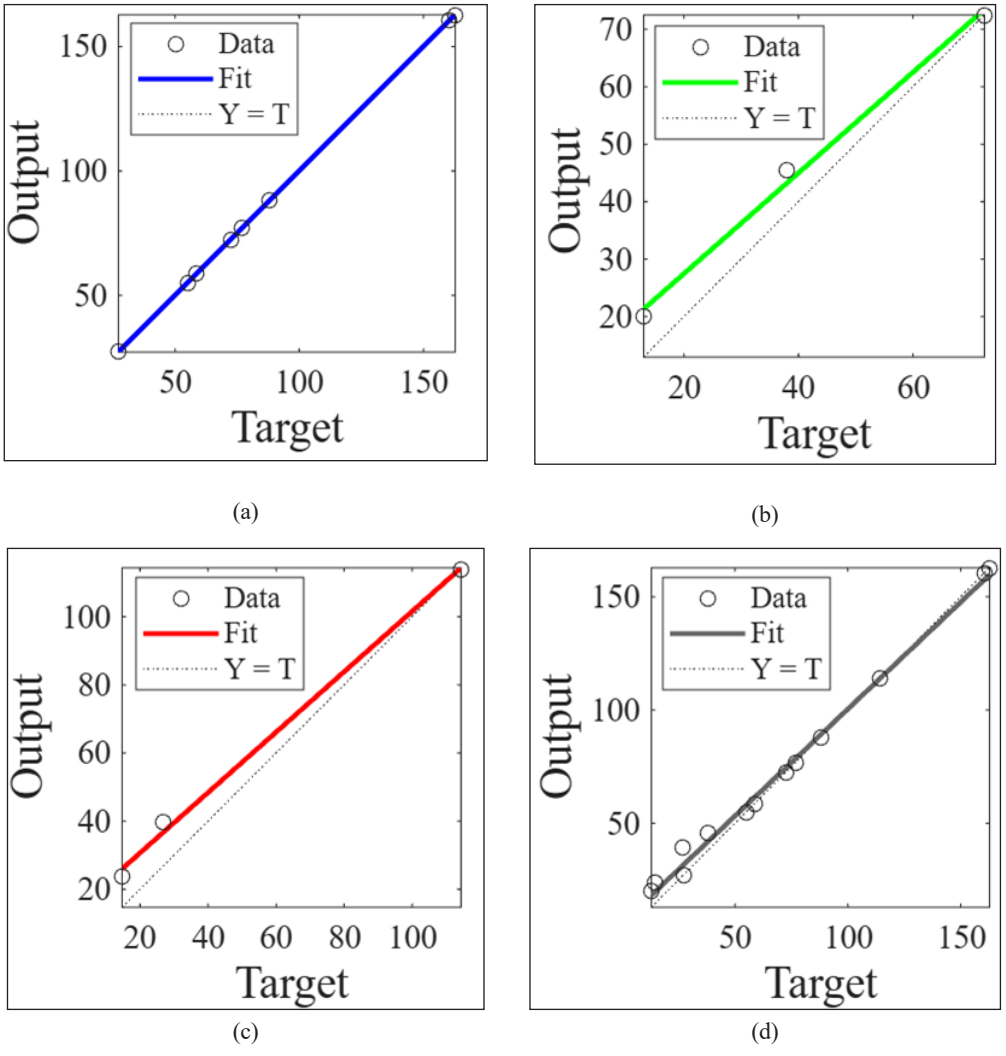


Figure 9. The results for regression analysis: (a) training, (b) validation, (c) test, and (d) overall data
Note. The dotted line represents the ideal line where the output (Y) = target for a perfect fit (T)

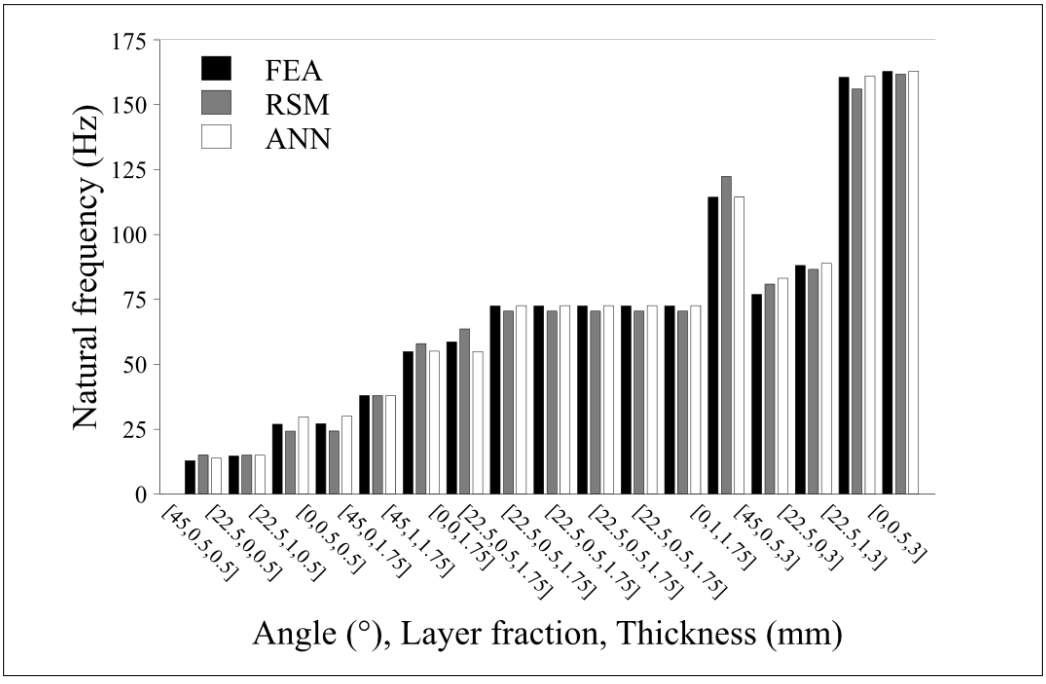


Figure 10. Comparison of natural frequency between the finite element analysis (FEA), response surface method (RSM), and artificial neural network (ANN)

of 0.9986, indicating accurate predictions, while the validation dataset shows a high *R* value of 0.99712, reflecting strong predictive performance. The model demonstrates accuracy and reliability, evidenced by a total *R* value of 0.99682 across the dataset. This study utilises a model summary that integrates Simulink with an ANN to predict natural frequencies. The utilised algorithms comprise the LM training algorithm, random data partitioning, and performance evaluation based on MSE. The accuracy of the ANN model in predicting natural frequencies is demonstrated in Table 4, with percentage errors typically remaining below 10.78%. The increased accuracy demonstrates the model's reliability, as shown in Figure 10. The model demonstrates accuracy, as most samples show errors below 1%. The predicts produced by ANSYS and ANN demonstrate considerable consistency. This agreement, which closely corresponds with the ANSYS results, indicates that the ANN accurately predicts natural frequencies.

Figure 10 compares the predicted natural frequencies obtained through two distinct methods. The predictions are predominantly aligned, with FEA's estimates marginally exceeding those of RSM. The intricate correlation between natural frequency and laminate factors, angle orientation, layer fraction, and plate thickness is emphasised by the initial peak in the bar graph, followed by a decline and subsequent increase in values. This pattern was arranged by increasing plate thickness, followed by layer fraction and angle

orientation. However, the initial spike offers critical insights for structural analysis and design optimisation by identifying factor combinations that significantly influence the stiffness and dynamic behaviour of the laminate. Thus, the initial spike indicates that low-thickness, completely carbon has provided high natural frequencies at 0°. The results of the present study align well with past studies (Njim et al., 2021; Rosaidi et al., 2023; Subrahmanyam et al., 2021).

CONCLUSION

This study illustrates the predictive abilities of ANN and RSM in examining the behaviour of hybrid composite laminates subjected to free vibration conditions. The models created through these methods exhibit differing levels of accuracy in predicting the system's responses, with the ANN model typically surpassing RSM regarding prediction accuracy. The accuracy of prediction is assessed using percentage error, offering a quantitative measure of the deviation between predicted and experimental results.

i. The ANN model demonstrated a low average percentage error of 2.81%, reflecting its effectiveness in identifying the underlying patterns within the data. Conversely, RSM exhibited a higher average percentage error of 4.96%. However, it still yielded valuable insights into the system's behaviour and functioned as a useful tool for preliminary analysis.

ii. The study identified that plate thickness significantly influences the free vibration of hybrid composite laminates, with layer fraction following as the next most impactful factor.

iii. This finding is essential for optimising laminate design for applications, guaranteeing the desired performance characteristics, including stability and resonance frequency.

iv. ANN demonstrate significant efficacy in managing complex, non-linear relationships and acquiring knowledge from extensive datasets, establishing itself as a formidable instrument in this field.

v. RSM provides a less complex and more straightforward approach, proving effective in situations where a simpler model suffices. The capacity to adjust to complex interactions among various variables renders it a suitable method for predicting the behaviour of composite materials in vibration analysis.

Both ANN and RSM significantly contribute to the study of hybrid composite laminates, each exhibiting distinct strengths. ANN demonstrate superior accuracy and flexibility, whereas RSM offers a dependable approach for preliminary evaluations and the optimisation of design factors. This study's findings emphasise the necessity of choosing

a suitable modelling technique according to the problem's complexity and the accuracy needed for the intended application. Future studies may concentrate on enhancing model calibration, investigating hybrid methodologies, or incorporating real-time data to achieve more reliable predictions.

ACKNOWLEDGEMENTS

The study has been funded by the Ministry of Higher Education of Malaysia under the Fundamental Research Grant Scheme (Grant no.: FRGS/1/2023/TK10/UITM/02/8).

REFERENCES

Alam, M. A., Ya, H. H., Azeem, M., Hussain, P. B., Salit, M. S., Khan, R., Arif, S., & Ansari, A. H. (2020). Modelling and optimisation of hardness behaviour of sintered Al/SiC composites using RSM and ANN: A comparative study. *Journal of Materials Research and Technology*, *9*(6), 14036–14050. https://doi.org/10.1016/j.jmrt.2020.09.087

An, H., Chen, S., & Huang, H. (2019). Multi-objective optimal design of hybrid composite laminates for minimum cost and maximum fundamental frequency and frequency gaps. *Composite Structures*, *209*, 268–276. https://doi.org/10.1016/j.compstruct.2018.10.075

Antony, J. (2014). *Design of experiments for engineers and scientists* (2nd ed.). Elsevier. https://doi.org/10.1016/C2012-0-03558-2

Arunachalam, S. J., Saravanan, R., Sathish, T., Giri, J., & Kanan, M. (2024). Mechanical assessment for enhancing hybrid composite performance through silane treatment using RSM and ANN. *Results in Engineering*, *24*, 103309. https://doi.org/10.1016/j.rineng.2024.103309

Arvinda Pandian, C. K., & Jailani, H. S. (2021). Formulation and characterisation of jute fabric–linen fabric–fumed silica–epoxy hybrid laminates. *Journal of the Textile Institute*, *113*(1), 151–158. https://doi.org/10.1080/00405000.2020.1866271

Bezerra, M. A., Santelli, R. E., Oliveira, E. P., Villar, L. S., & Escaleira, L. A. (2008). Response surface methodology (RSM) as a tool for optimization in analytical chemistry. *Talanta*, *76*(5), 965–977. https://doi.org/10.1016/j.talanta.2008.05.019

Boukarma, L., Aboussabek, A., El Aroussi, F., Zerbet, M., Sinan, F., & Chiban, M. (2023). Insight into mechanism, Box-Behnken design, and artificial neural network of cationic dye biosorption by marine macroalgae *Fucus spiralis*. *Algal Research*, *76*, 103324. https://doi.org/10.1016/j.algal.2023.103324

Box, G. E. P., & Behnken, D. W. (1960). Some new three level designs for the study of quantitative variables. *Technometrics*, *2*(4), 455–475. https://doi.org/10.1080/00401706.1960.10489912

Bulut, M., Erkliğ, A., & Yeter, E. (2016). Experimental investigation on influence of Kevlar fiber hybridization on tensile and damping response of Kevlar/glass/epoxy resin composite laminates. *Journal of Composite Materials*, *50*(14), 1875–1886. https://doi.org/10.1177/0021998315597552

Chisena, R. S., Chen, L., & Shih, A. J. (2021). Finite element composite simplification modeling and design of the material extrusion wave infill for thin-walled structures. *International Journal of Mechanical Sciences*, *196*, 106–276. https://doi.org/10.1016/j.ijmecsci.2021.106276

Crawley, E. F. (1979). The natural modes of graphite/epoxy cantilever plates and shells. *Journal of Composite Materials*, *13*(3), 195–205. https://doi.org/10.1177/002199837901300302

de Paula Ferreira, W., Armellini, F., & De Santa-Eulalia, L. A. (2020). Simulation in industry 4.0: A state-of-the-art review. *Computers and Industrial Engineering*, *149*, 106868. https://doi.org/10.1016/j.cie.2020.106868

Erkli̇, A., Bulut, M., & Yeter, E. (2015). The effect of hybridization and boundary conditions on damping and free vibration of composite plates. *Science and Engineering of Composite Materials*, *22*(5), 565–571. https://doi.org/10.1515/secm-2014-0070

Jadee, K. J., Abed, B. H., & Battawi, A. A. (2020). Free vibration of isotropic plates with various cutout configurations using finite elements and design of experiments. In *IOP Conference Series: Materials Science and Engineering* (Vol. 745, No. 1, p. 012080). IOP Publishing. https://doi.org/10.1088/1757-899X/745/1/012080

Kallannavar, V., Kumaran, B., & Kattimani, S. C. (2020). Effect of temperature and moisture on free vibration characteristics of skew laminated hybrid composite and sandwich plates. *Thin-Walled Structures*, *157*, 107113. https://doi.org/10.1016/j.tws.2020.107113

Kumar C, H., & Swamy, R. P. (2021). Fatigue life prediction of glass fiber reinforced epoxy composites using artificial neural networks. *Composites Communications*, *26*, 100812. https://doi.org/10.1016/j.coco.2021.100812

Leissa, A. W. (1973). The free vibration of rectangular plates. *Journal of Sound and Vibration*, *31*(3), 257–293. https://doi.org/10.1016/S0022-460X(73)80371-2

Li, M., Soares, C. G., Liu, Z., & Zhang, P. (2024). Free and forced vibration analysis of carbon/glass hybrid composite laminated plates under arbitrary boundary conditions. *Applied Composite Materials*, *31*, 1687–1710. https://doi.org/10.1007/s10443-024-10235-y

Ma, J., & Gao, H. (2024). Free vibration analysis of hybrid laminated thin-walled cylindrical shells containing multilayer FG-CNTRC plies. *Acta Mechanica*, *235*, 7711–7731. https://doi.org/10.1007/s00707-024-04082-y

Malik, M. H., & Arif, A. F. M. (2013). ANN prediction model for composite plates against low velocity impact loads using finite element analysis. *Composite Structures*, *101*, 290–300. https://doi.org/10.1016/j.compstruct.2013.02.020

Mallick, A., Mandal, R., Mondal, N., Sarkar, S., Biswas, N., Maji, B., & Majumdar, G. (2024). Parametric optimization and minimization of corrosion rate of electroless Ni–P coating using Box-Behnken design and Artificial Neural Network. *Results in Surfaces and Interfaces*, *15*, 100228. https://doi.org/10.1016/j.rsurfi.2024.100228

Mat Jusoh, S., Zulkifli, M. F. R., Abdullah, S., Shafi, A. R., Mohd Nasir, M. F., Maulana, U. N., Mohd Radzi, F. S., & Zainulabidin, S. H. (2024). Response Surface Methodology for optimisation of factor on water absorption of natural fibre hybrid composite in boat construction. *Pertanika Journal of Science and Technology*, *32*(S5), 123–132. https://doi.org/10.47836/pjst.32.S5.06

Mengistu, A., Andualem, G., Abewaa, M., & Birhane, D. (2024). Keratin extraction optimization from poultry feather using response surface-Box-Behnken experimental design method. *Results in Engineering*, *22*, 102360. https://doi.org/10.1016/j.rineng.2024.102360

Merzuki, M. N. M., Rejab, M. R. M., Sani, M. S. M., Zhang, B., & Quanjin, M. (2019). Experimental investigation of free vibration analysis on fibre metal composite laminates. *Journal of Mechanical Engineering and Sciences*, *13*(4), 5753–5763.

Mirziyod, M., Ibrokhimovich, S. I., & Khudoyberdievich, T. M. (2019). Dynamics of structural-inhomogeneous laminate and shell mechanical systems with point constraints and focused masses. Part 2. Statement of the problem of forced oscillations, methods of solution, computational algorithm and numerical results. *Journal of Applied Mathematics and Physics*, *7*(11), 2671–2684. https://doi.org/10.4236/jamp.2019.711182

Müzel, S. D., Bonhin, E. P., Guimarães, N. M., & Guidi, E. S. (2020). Application of the finite element method in the analysis of composite materials: A review. *Polymers*, *12*(4), 818. https://doi.org/10.3390/POLYM12040818

Njim, E. K., Bakhy, S. H., & Al-Waily, M. (2021). Analytical and numerical investigation of free vibration behavior for sandwich plate with functionally graded porous metal core. *Pertanika Journal of Science and Technology*, *29*(3), 1655–1682. https://doi.org/10.47836/pjst.29.3.39

Norman, M. A. M., & Mahmud, J. (2021). Investigation of natural frequencies on Kevlar/glass hybrid laminated composite plates using finite element simulation. *International Transaction Journal of Engineering, Management, and Applied Sciences and Technologies*, *12*(9), 1–8. https://doi.org/10.14456/ITJEMAST.2021.169

Norman, M. A. M., Sivakumar, K., Ab Patar, M. N. A., & Mahmud, J. (2022). A prediction model for natural frequencies on Kevlar/glass hybrid laminated composite using Artificial Neural Networks (ANN). *International Journal of Integrated Engineering*, *14*(5), 207–214. https://doi.org/10.30880/ijie.2022.14.05.023

Norman M. A. M., Zainuddin M. A., & Jamaluddin, M. (2018). The effect of various fiber orientations and boundary conditions on natural frequencies of laminated composite Beam. *International Journal of Engineering and Technology*, *7*(3.11), 67–71. https://doi.org/10.14419/ijet.v7i3.11.15932

Polat, S., & Sayan, P. (2019). Application of response surface methodology with a Box–Behnken design for struvite precipitation. *Advanced Powder Technology*, *30*(10), 2396–2407. https://doi.org/10.1016/j.apt.2019.07.022

Pratap, B., Mondal, S., & Rao, B. H. (2024). Prediction of compressive strength of bauxite residue-based geopolymer mortar as pavement composite materials: An integrated ANN and RSM approach. *Asian Journal of Civil Engineering*, *25*, 597–607. https://doi.org/10.1007/s42107-023-00797-w

Pushparaj, P., & Suresha, B. (2016). Free vibration analysis of laminated composite plates using finite element method. *Polymers and Polymer Composites*, *24*(7), 529–538. https://doi.org/10.1177/096739111602400712

Quanjin, M., Merzuki, M. N. M., Rejab, M. R. M., Sani, M. S. M., & Zhang, B. (2022). Numerical investigation on free vibration analysis of Kevlar/glass/epoxy resin hybrid composite laminates. *Malaysian Journal on Composites Science and Manufacturing*, *9*(1), 11–21. https://doi.org/10.37934/mjcsm.9.1.1121

Ratnaparkhi, S. U., & Sarnobat, S. S. (2013). Vibration analysis of composite plate. *International Journal of Modern Engineering Research*, *3*(1), 377–380.

Reddy, B. S., Reddy, M. R. S., & Reddy, V. N. (2013). Vibration analysis of laminated composite plates using design of experiments approach. *International Journal of Scientific Engineering and Technology*, *2*(1), 40–49.

Reddy, M. R. S., Reddy, B. S., Reddy, V. N., & Sreenivasulu, S. (2012). Prediction of natural frequency of laminated composite plates using artificial neural networks. *Engineering*, *4*(6), 329–337. https://doi.org/10.4236/eng.2012.46043

Rosaidi, M. H. M., Mahmud, J., Abdul Rahman, S. M., Abdul Halim, N. H., Raja Abdullah, R. I., & Ab Patar, M. N. A. (2023). Natural frequencies optimisation of hybrid composite laminates using Response Surface Method. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *31*(2), 197–209. https://doi.org/10.37934/araset.31.2.197209

Safri, S. N. A., Sultan, M. T. H., Jawaid, M., & Jayakrishna, K. (2018). Impact behaviour of hybrid composites for structural applications: A review. *Composites Part B Engineering*, *133*, 112–121. https://doi.org/10.1016/j.compositesb.2017.09.008

Sai Vivek, K. (2016). Free vibration of skew laminated composite plates with circular cutout by finite element method. *International Journal of Modern Engineering Research*, *6*(6), 15–23.

Samsudin, A. H., & Mahmud, J. (2015). The effect of lamination scheme and angle variations to the displacements and failure behaviour of composite laminate. *Journal Technology Science Engineering*, *76*(11), 113–121. https://doi.org/10.11113/jt.v76.5921

Saravanakumar, S., Sathiyamurthy, S., Pathmanaban, P., & Devi, P. (2024). Integrating machine learning and response surface methodology for analyzing anisotropic mechanical properties of biocomposites. *Composite Interfaces*, *31*(1), 1–28. https://doi.org/10.1080/09276440.2023.2260239

Sathishkumar, N., Selvam, R., Kumar, K. M., Abishini, A. H., Rahman, T. K., & Mohanaranga, S. (2022). Influence of garnet abrasive in drilling of Basalt–Kevlar–Glass fiber reinforced polymer cross ply laminate by Abrasive Water Jet Machining process. *Materials Today: Proceedings*, *62*(Part 2), 1361–1368. https://doi.org/10.1016/j.matpr.2022.04.861

Shi, D., He, D., Wang, Q., Ma, C., & Shu, H. (2020). Free vibration analysis of closed moderately thick cross-ply composite laminated cylindrical shell with arbitrary boundary conditions. *Materials*, *13*(4), 884. https://doi.org/10.3390/ma13040884

Shokuhfar, A., Khalili, S. M. R., Ashenai Ghasemi, F., Malekzadeh, K., & Raissi, S. (2008). Analysis and optimization of smart hybrid composite plates subjected to low-velocity impact using the response surface methodology (RSM). *Thin-Walled Structures*, *46*(11), 1204–1212. https://doi.org/10.1016/j.tws.2008.02.007

Sreekanth, T. G., Senthilkumar, M., & Reddy, S. M. (2023). Artificial neural network based delamination prediction in composite plates using vibration signals. *Fracture and Structural Integrity*, *17*(63), 37–45. https://doi.org/10.3221/IGF-ESIS.63.04

Subrahmanyam, A. P. S. V. R., Srinivasa Rao, P., & Siva Prasad, K. (2021). Enhancement of surface quality of DMLS aluminium alloy using RSM optimization and ANN modelling. *Journal of Mechanical Engineering*, *18*(3), 37–56.

Swamy, G. J., Sangamithra, A., & Chandrasekar, V. (2014). Response surface modeling and process optimization of aqueous extraction of natural pigments from *Beta vulgaris* using Box-Behnken design of experiments. *Dyes and Pigments*, *111*, 64–74. https://doi.org/10.1016/j.dyepig.2014.05.028

Triefenbach, F. (2008). *Design of experiments: The D-optimal approach and its implementation as a computer algorithm* [Unpublished Bachelor's dissertation]. South Westphalia University of Applied Sciences.

Warburton, G. B. (1954). The vibration of rectangular plates. *Proceedings of the Institution of Mechanical Engineers*, *168*(1), 371–384. https://doi.org/10.1243/pime_proc_1954_168_040_02

Pertanika J. Sci. & Technol. 33 (6): 2615 - 2639 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Case Study

Mid-to-Post-Pandemic Mobility Patterns: A Case Study of Malaysia Highlighting Differential Trends in Work and Non-Work Trips

Megat-Usamah Megat-Johari[1]*, Nusayba Megat-Johari[2], Nischal Gupta[3], Hisham Jashami[4] and Teik Hua Law[1]

[1]Department of Civil Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[2]Department of Road Engineering and Environment, Malaysian Institute of Road Safety Research, 43000 Kajang, Selangor, Malaysia
[3]Department of Civil and Environmental Engineering, Michigan State University, East Lansing, Michigan 48824, United States
[4]Department of Civil and Construction Engineering, Oregon State University, Corvallis, Oregon 97331, United States

ABSTRACT

The coronavirus disease 2019 (COVID-19) pandemic has had a significant impact on people's mobility patterns worldwide, with some effects still prevalent. This study examines mobility trends in Malaysia before, during, and after the COVID-19 pandemic period. Using trip information from community mobility reports, along with data on the number of COVID-19 cases and vaccination rates, this study investigates how work and non-work trips changed during 2020-2022 while also controlling for several socio-demographic factors. The results showed a strong correlation between mobility and the severity of lockdown, with a significant decline in all trip types during lockdowns. The number of new COVID-19 cases and deaths was also found to impact people's mobility. When controlled for other factors, vaccination rates, however, did not seem to affect trip patterns. The results also showed a rebound in all trip types during the post-pandemic period (April 2022 to October 2022); however, the recovery rates differed, with non-work trips recovering more rapidly than work trips. The study highlights the immediate impact of the post-pandemic on work habits and public behaviour and discusses implications for policymakers in general.

Keywords: COVID-19, mobility, non-work trips, post-pandemic, stay-at-home order, transit trips, work from home, work trips

ARTICLE INFO

Article history:
Received: 02 December 2024
Accepted: 29 April 2025
Published: 08 October 2025

DOI: https://doi.org/10.47836/pjst.33.6.10

E-mail addresses:
megatusamah@upm.edu.my (Megat-Usamah Megat-Johari)
nusayba@miros.gov.my (Nusayba Megat-Johari)
guptanis@msu.edu (Nischal Gupta)
hisham.jashami@oregonstate.edu (Hisham Jashami)
lawteik@upm.edu.my (Teik Hua Law)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

BACKGROUND

The COVID-19 pandemic brought a myriad of challenges to societies worldwide. One of the fundamental aspects profoundly affected was the mobility of individuals, as lockdowns and restrictions swept across the globe. It is estimated that over a third of the global population has faced significant mobility restrictions since the beginning of the COVID-19 pandemic (International Organization of Migration [IOM], n.d.). Limiting population movement stands out as a key strategy employed by various countries, including Malaysia (Dias et al., 2021; Shibayama et al., 2021). During these lockdowns, the public was urged to travel only for essential activities as an effective measure for disease control (Badr et al., 2020; Bergman & Fishman, 2023). A systematic review of more than 40 studies reporting on COVID-19 within the first half of 2020, investigating the correlation between human mobility and the transmission of COVID-19, revealed that while limiting human mobility decreased the spread of the virus, the efficacy and strictness of policy enforcement exhibited temporal and spatial variations during different phases of the pandemic (Zhang et al., 2022). Mobility patterns were significantly affected during the pandemic compared to pre-pandemic times due to COVID-19 and related sanitary concerns about overcrowding on certain modes of transportation. Planned transportation post-COVID indicated that challenges may persist for buying into public transport use compared to pre-COVID (Balbontin et al., 2024; Mukhtar et al., 2024).

Studies have been conducted on the effect of COVID-19 on mobility, as shifts in work-related commuting were observed, and working from home was provided as an alternative to commuting. A survey in Australia looked at the effect of the pandemic on respondents' travel behaviour by employment status, area of residence, and work-from-home (WFH) factor. The number of commuting and work-related trips reduced in metropolitan and regional areas, while non-work-related trips increased (Beck & Hensher, 2020a). In another study, post-pandemic work schedules of tech companies in the San Francisco Bay Area indicated a 180-degree shift to pre-pandemic behaviour. The post-pandemic period saw more than 95% of workers working either remotely or via a hybrid format, while almost three-quarters of the workers commuted to the office daily, pre-pandemic (Tan et al., 2023).

Public transport ridership was reported to be significantly reduced in many countries, between 50 to 80% during the early stages of the pandemic (Gramsch et al., 2022; Haseeb & Mitra, 2024; Kar et al., 2022). Changes in mobility behaviour during the pandemic, as well as in its aftermath, have also been attributed to the perception of users. A study conducted in Poland focused on public transport users found that the future of public transport in urban areas, as well as passengers' willingness to utilize it post-epidemic, hinges on the perceived comfort and safety experienced during the epidemic (Przybylowski et al., 2021). Initial predictions had indicated that public transport use would drop sharply during the pandemic, but was expected to recover post-pandemic, though not fully, remaining

below the pre-pandemic level (Currie et al., 2021). A study in Spain indicated that with necessary countermeasures in place, respondents were willing to keep using public transport post-pandemic (Awad-Núñez et al., 2021). While a study on three phases: pre-pandemic, during pandemic, and post-pandemic, found that modality profiles significantly changed during the pandemic but are expected to return to the pre-pandemic condition with time, especially private vehicle use and active mobility (Lodi et al., 2024; Ong et al., 2024). The improvement in public transport usage post-pandemic was reported in various countries, where transit use rebounded as restrictions were progressively eased (Bonera & Martinelli, 2023). However, in some areas, the shift from public transport to private vehicles observed during the pandemic's restrictive periods has persisted (Department for Transport [DfT], 2023; Iglesias & Raveau, 2024).

During the pandemic, travel related to non-essential services was reduced, and remote working was expected. Thus, work-related travel during the pandemic, especially during lockdowns, saw drastic reductions. The effect on work-travel, however, has been shown to differ between different socioeconomic statuses (SES) (Chang et al., 2021; Matson et al., 2021). High SES was found to be associated with a significant reduction in work travel and participation in recreational travel, while the medium and low SES groups travelled more for work, indicative of fewer opportunities to work from home as compared to the high SES group (Politis et al., 2021; Xi et al., 2023).

Similar to work-related trips, studies indicated that income was identified as a significant factor influencing the frequency of essential non-work-related trips as well (Beck & Hensher, 2020b; Sharma et al., 2020). High-income individuals were found to be less prone to completely avoiding travel during the transition to the lockdown period compared to those in the low-income group (Pawar et al., 2021). The frequency of travel for essential trips was notably higher than for non-essential ones. This suggests that people prioritized essential activities like grocery shopping during the transition to the lockdown period while minimizing recreational trips (Pawar et al., 2021). However, respondents in low-income families with children reported that the frequency of grocery shopping decreased during the pandemic (Sharma et al., 2020).

Although an extensive number of studies have evaluated the impact of the COVID-19 pandemic on people's movements, the effects during the post-pandemic period remain somewhat understudied, particularly in Malaysia, where empirical evaluations are still limited. Even though the lockdowns have been lifted throughout the country, their effects on mobility persist. The pandemic prompted significant shifts in work habits, transportation choices, and public activity patterns, some of which may have long-term implications. Understanding these changes is crucial for transportation planning, public transit management, and policymaking, particularly in preparing for future disruptions. The primary objective of this research is to investigate changes in mobility patterns during

and post-pandemic, exploring the impact of the restrictions imposed on various aspects of social and economic structure. To that end, mixed-effects linear regression models are estimated to discern the impacts of various lockdown policies and other important socio-economic and COVID-19-related information on various types of trips, including work trips, trips to transit stations, retail and recreational trips, and trips to grocery stores and pharmacies. Furthermore, as Malaysia declared COVID-19 an endemic issue in April 2022, signifying a transformative juncture in the country's approach, this study also investigates the subsequent recovery of mobility patterns. Beyond merely quantifying movements, this research seeks to unravel the socio-economic factors underlying these shifts, providing a comprehensive understanding of how a nation adapts and rebounds in the aftermath of a global health crisis.

DATA DESCRIPTION

This study gathered various data from different sources to understand how mobility changed during the pandemic and how it progressed after Malaysia declared COVID-19 an endemic issue in 2022. The following subsections briefly describe each data source utilized in this study.

COVID-19 Google Mobility Reports

The study leveraged open-access data from Google on mobility reports in Malaysia corresponding to the COVID-19 pandemic (Google, 2022). The data provides information on how movement trends to different places changed in each geographic region. It shows how visits and time spent at different locations changed with respect to a baseline value. The daily changes in movement are compared to a baseline value for the same day of the week (Equation 1), which, in this case, was the median value of five weeks from January 3 to February 6, 2020 (i.e., the pre-pandemic period). The dataset provided six different categories of places that are important to social distancing efforts, in addition to access to essential services. These categories included workplace, residential area, retail and recreation (e.g., restaurants and shopping centres), transit stations, grocery and pharmacy (i.e., these are combined as they can be considered as essential trips), and parks. For the purposes of this study, four categories were assessed: workplace transit station, retail and recreation, and grocery and pharmacy.

$$\%\ change\ in\ movement = \frac{Daily\ mobility\ -Median\ value\ of\ mobility\ \ of\ the\ same\ day\ from\ Jan\ 3\ to\ Feb\ 6{,}2020}{Median\ value\ of\ mobility\ \ of\ the\ same\ day\ from\ Jan\ 3\ to\ Feb\ 6{,}2020} \times 100\% \quad [1]$$

In this study, a three-year period of data, dating back from February 2020 to October 2022, was utilized. The end period was selected because it was the last month for which the data was updated. The data was arranged at the state-day level, with each observation (or row) representing the percentage change in movement for the state on a particular day. In the preliminary analysis, a total of 15,579 state-day observations were considered.

COVID-19 Data

The Ministry of Health provides an open-access database for COVID-19 information for each state and federal territory in Malaysia (Ministry of Health Malaysia [MOH], n.d.). This weekly updated data consists of information on the number of daily COVID-19 cases, COVID-19 deaths, and vaccination rates. These data were obtained for the time period that coincides with the COVID-19 community mobility reports.

Demographic Background

Data on demographic background and socioeconomic status were obtained from the official website of the Department of Statistics Malaysia (DOSM). These data consist of information on population demographics (by age, sex, development area, and race), income, and poverty rate (i.e., percentage of population living under the national poverty line income, US$471), among others. It should be noted that some of this information (e.g., poverty and unemployment rate) is at the state-month level, whereas other data, such as population and median income, are at the state-year level.

COVID-19 Timeline Data

The Malaysian National Security Council (NSC) provided information to the country on the standard operating procedure (SOP) with the help of other government sectors on managing the COVID-19 pandemic. One of the crucial pieces of information that the agency distributed to the public during the pandemic period was the lockdown information. In Malaysia, the lockdown due to COVID-19 can be divided into three different categories: Movement Control Order (MCO), Conditional MCO (CMCO), and Recovery MCO (RMCO). The main difference between these three categories is the level of restriction imposed on the public within their geographical region, where MCO had the strictest lockdown, and RMCO had the least restriction on people's movements.

Figure 1 shows the COVID-19 timeline in Malaysia. After the World Health Organization (WHO) declared the COVID-19 outbreak a global pandemic on March 11, 2020, the Malaysian government announced its first MCO on March 18, 2020 (Tang, 2020), which lasted until May 3, 2020. During this period, many restrictions were imposed, which included an interstate and inter-district travel ban, closure of all houses of worship

and most business premises, a maximum of 10 km travel radius, closure of all educational institutions, shutting down of all government and private sectors except those involved with daily essential services, border closure, and gathering bans, among others. To ensure compliance from the public with these restrictions, road closures were introduced by law enforcement throughout the country.

Beginning May 4, 2020, Malaysia transitioned into CMCO. Most economic sectors could operate by following the strict SOP introduced by the NSC. Houses of worship and some business premises were allowed to operate at limited capacity. However, educational institutions remain closed, and interstate travel is still banned. As COVID-19 cases continued to drop, the government started to further loosen the restrictions by introducing RMCO throughout the country starting June 10, 2020. During this period, interstate travel was allowed, tourism started to open, gathering activities with limited capacity were permitted, and educational institutions started to reopen in stages.

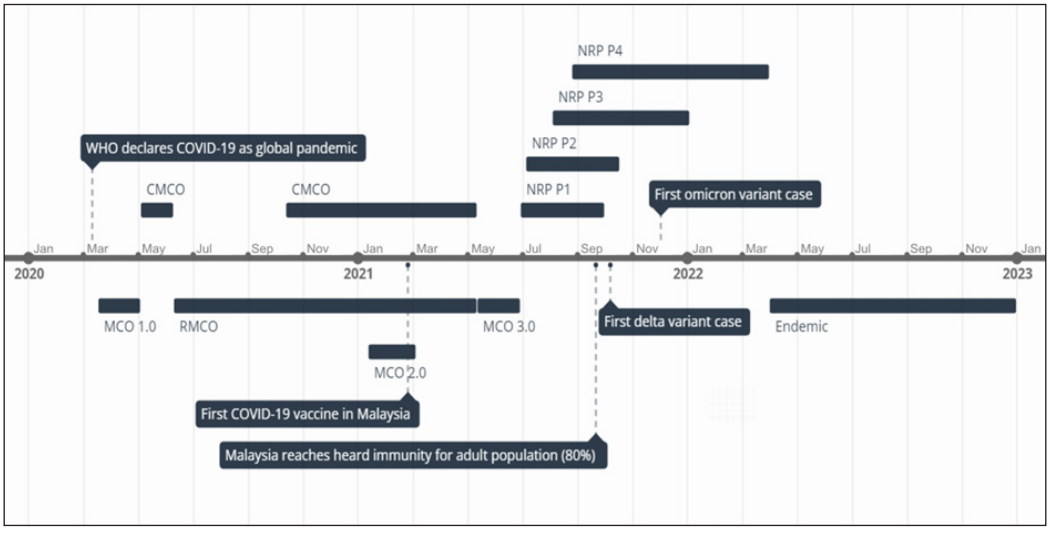


Figure 1. Timeline of lockdown in Malaysia
Note. WHO = World Health Organization; MCO = Movement Control Order; CMCO = Conditional Movement Control Order; RMCO = Recovery Movement Control Order; NRP = National Recovery Plan

As the restrictions continued to loosen up, the number of COVID-19 cases started to increase in several localities, resulting in the reinstatement of CMCO in these areas. As time progressed, more states started to transition back into CMCO, and by December 2020, 12 out of 16 localities were placed under CMCO. As the situation started to worsen, a second MCO was reintroduced between January 13 and March 4, 2021 (some states started early). After March 4, 2021, some states transitioned back to CMCO and some to RMCO, before the third MCO was reinstated for the whole nation on May 12, 2021. The

third MCO lasted until June 28, 2021, before the government introduced a new plan to combat COVID-19 called the National Recovery Plan (NRP).

The NRP consisted of four phases, and the transition between each phase depended on the average daily cases of COVID-19, the capacity of the public health system (i.e., rate of intensive care unit ward occupancy), and the percentage of the population that was fully vaccinated. Note that phase 1 of NRP was similar to the MCO restriction measures. The NRP-1 spanned from June to September 2021. For the NRP-2, it started from early July 2021 until October 2021, with some restrictions lifted for complete vaccination recipients. The NRP-3 period began in August 2021 and ended in January 2022, with inter-district travel allowed regardless of vaccination status. The last phase of NRP started in August 2021, with social activities allowed with limited capacity, and the interstate travel ban was lifted for fully vaccinated individuals. On April 1, 2022, Malaysia transitioned into the endemic phase, where all restrictions that were previously introduced were completely removed.

Data Summary

Table 1 shows the descriptive statistics of key variables of interest in this study. This includes the percentage change in mobility, socioeconomic and demographic factors, the COVID-19 time period, COVID-19 cases, and vaccination status. It should be noted that there are several variables obtained based on state-day level information, such as the percentage change in mobility, new COVID-19 cases and deaths, vaccination, and lockdown status. The remaining variables are based on the state-month level information, e.g., the unemployment rate information varies from state to state but remains the same within each month.

Table 1
Descriptive statistics of variables of interest

Variable	Min.	Max.	Mean	Std. Dev.
Sociodemographic factors				
Median income in RM1,000	3.01	10.23	5.73	2.01
Percentage of population below the national poverty line	0.10	25.30	7.58	6.02
Percentage of unemployment	1.00	8.00	4.04	1.58
Percentage of the Malay population	6.93	95.01	59.52	25.11
Percentage of the Chinese population	0.51	41.28	17.36	12.11
Percentage of the Indian population	0.17	13.49	4.88	4.36
Percentage of the rural population	0.00	55.90	26.04	18.02
Percentage of the female population	46.58	50.92	48.34	1.16
Timeline				

Table 1 *(continue)*

Variable	Min.	Max.	Mean	Std. Dev.
Sociodemographic factors				
Pre-pandemic period (Feb. 15 - Mar. 10, 2020)	0	1	0.03	0.16
No lockdown (Mar. 11 - Mar. 17, 2020)	0	1	0.01	0.08
MCO-1, (Mar. 18 - May 03, 2020)	0	1	0.05	0.21
MCO-2, (Jan. 13 – March 4, 2021)	0	1	0.04	0.19
MCO-3, (May 12 - Jun. 28, 2021)	0	1	0.05	0.22
CMCO	0	1	0.12	0.32
RMCO	0	1	0.23	0.42
NRP-1 (Jun. 29 - Sep. 30, 2021)	0	1	0.04	0.19
NRP-2 (Jul. 05 - Oct. 17, 2021)	0	1	0.05	0.22
NRP-3 (Aug. 04 - Dec. 31, 2021)	0	1	0.04	0.19
NRP-4 (Aug. 26, 2021 - Mar. 31, 2022)	0	1	0.16	0.36
No lockdown (Apr. 1 - Oct. 31, 2022)	0	1	0.20	0.40
COVID-19 cases and vaccination				
New COVID-19 cases per day per 100,000 people (7-day moving average)	**0.00**	**391.57**	**15.40**	**27.64**
New COVID-19 death per day per 100,000 people (7-day moving average)	0.00	4.32	0.10	0.28
Percentage of fully vaccinated people	0.00	159.47	37.63	43.86
Percentage change in mobility				
Workplace	-92	21	-18.24	19.16
Transit station	-100	247	-30.19	35.95
Retail and recreation	-90	165	-19.85	27.14
Grocery and pharmacy	-74	188	7.43	25.32

Note. Min. = Minimum; Max. = Maximum; Std. Dev. = Standard deviation; MCO = Movement Control Order; CMCO = Conditional Movement Control Order; RMCO = Recovery Movement Control Order; NRP = National Recovery Plan

Based on Table 1, the average median household income across states was RM5,730 per month. Figure 2 shows the average median income level (monthly) by state. The state of Kelantan shows the lowest income level when compared to other states. In terms of the poverty rate, the average across the states was 7.6%. The worst rate recorded was during the 2020 pandemic year, when the percentage of the population living under the national poverty line was 25.3% (DOSM, 2021). The unemployment rate ranged from 1 to 8%, with an average of 4%. Similar to the poverty rate, the highest average recorded for the unemployment rate was during the 2020 pandemic year, when Malaysia saw approximately 5.3% unemployment in the month of May (DOSM, 2025). For the population by race, an average of 59.5% of people are Malay. The data also showed that 26% of people resided in rural areas, on average.

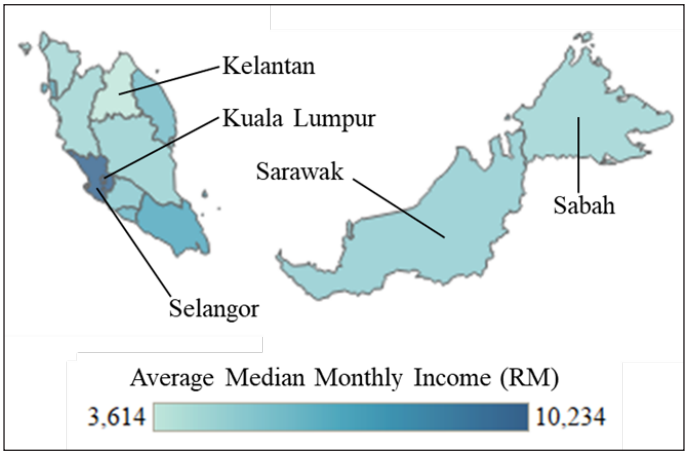


Figure 2. Average median monthly income by state

The timeline was divided into 12 categories based on the lockdowns. Note that the categories during the pandemic period correspond to the level of lockdown, and they reflect the restrictions imposed. Three percent of the data was from the pre-pandemic period; 20% from the post-pandemic period, and the remaining data was from the pandemic period.

The new COVID-19 cases and deaths were reported daily in Malaysia. Within the context of this study, a 7-day running average was calculated to account for any fluctuations in the reported data. Additionally, the data was normalized by the population of the state, resulting in the number of cases and deaths per 100,000 population. On average, new COVID-19 cases and deaths per 100,000 population across all states-days were 15.4 and 0.1, respectively. The vaccination rates represent the percentage of the population receiving at least two doses of the COVID-19 vaccine. On average, 38% of the people were fully vaccinated across states during the study period. Note that the maximum value recorded exceeds 100%. This may be due to some percentage of people receiving the vaccine from other states where they live. Particularly states that border federal territories, or vice versa.

In terms of mobility, trips to the workplace, transit stations, and retail and recreational trips reduced by 18, 30, and 20%, respectively, on average compared to January 2020 data; while grocery and pharmacy trips increased by 7%. Note that the data from Table 1 shows the aggregate information during the entire study period. Figure 3 shows the disaggregated information related to the percentage change in mobility by time and state. The red and blue colours indicate an increase and a decrease in the percentage change in mobility, respectively. The darker the colour, the greater the percentage change.

The pre-lockdown period showed only marginal changes in mobility for all trip types. During the lockdown periods, mobility for all trip types saw the highest reductions across the state due to the strictest lockdown imposed. As the period transitioned to the post-

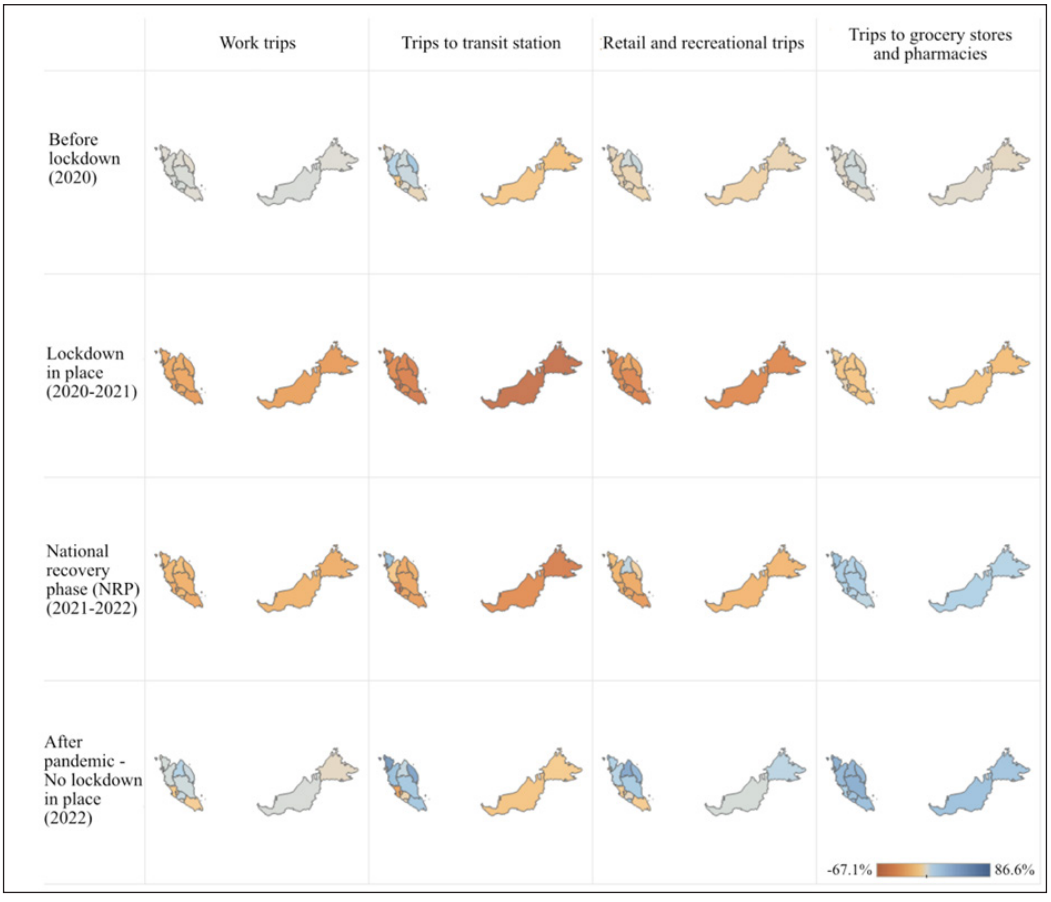


Figure 3. Average percentage change in mobility by time period and state

pandemic phase, all states showed an increase in mobility across all trip types. However, a few states and federal territories still experienced a negative change in mobility, particularly for work-related trips.

As time progressed, the percentage change in mobility started to increase. After the government announced the transition from global pandemic to endemic, all non-work trips saw an increase in mobility when compared to the pre-pandemic period. However, for workplace mobility, the average percentage change in mobility was still negative during the endemic period. Figure 4 shows the further disaggregation of the average daily trends of mobility for all trip types in Malaysia. It should be noted that the periods of holidays and weekends have not been removed from the data. These periods are indicated by sharp reductions in mobility, particularly for workplace trips, even in 2021-2022.

The global efforts to contain the pandemic resulted in the development of a vaccination against the virus. Figure 5 shows the trends in the number of COVID-19 cases and vaccines

administered in the country. It should be noted that even though COVID-19 was declared a pandemic in early March 2020 due to its rapid rise in some countries, such as the United States, the number of COVID cases in Malaysia remained substantially low during early 2020 due to the strict lockdowns in the early stages of the pandemic. As such, the rapid increase in the number of COVID-19 cases only began in October 2020, as seen in the figure. On February 24, 2021, the first COVID-19 vaccine was administered in Malaysia prior to its mass distribution (Anand, 2021). When comparing Figures 4 and 5, it is observed that as the percentage of the population receiving the full dosage increased, the percentage change in mobility also increased. By the end of the study period, more than 80% of Malaysia's population was fully vaccinated.

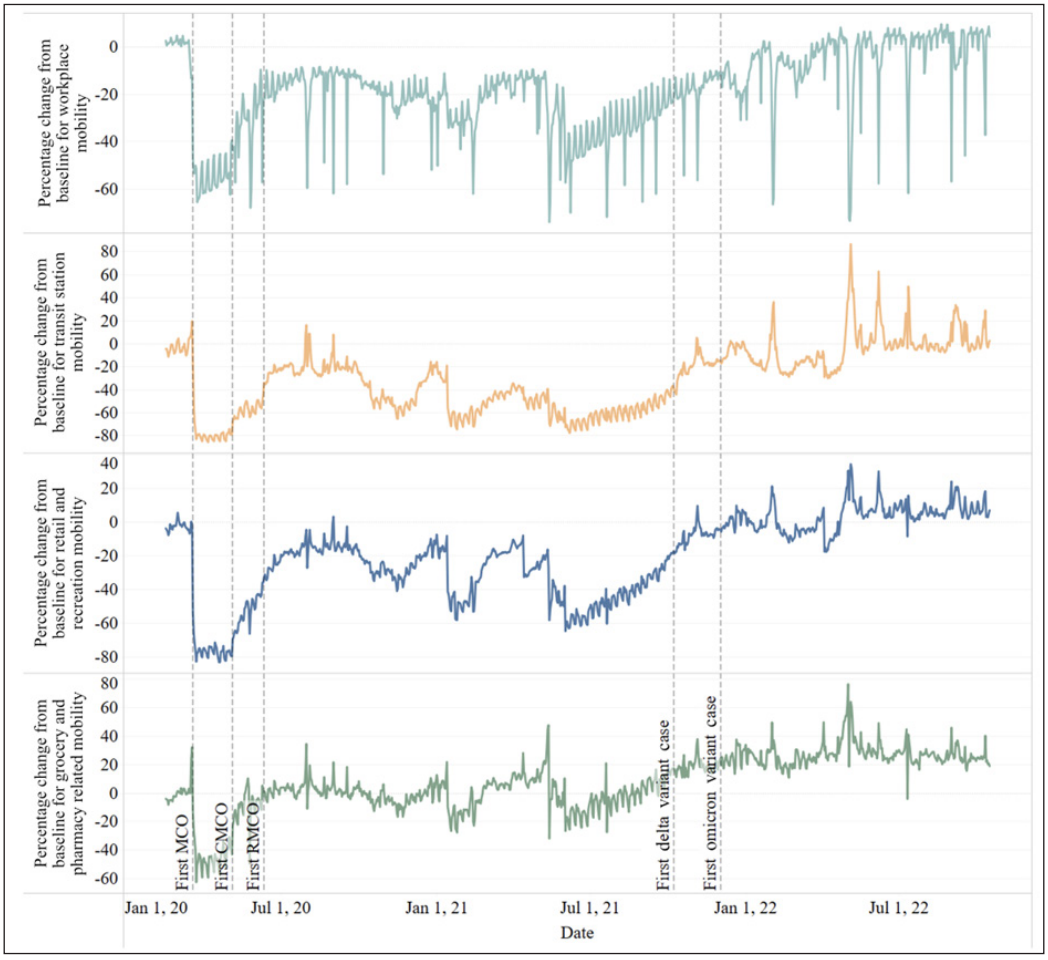


Figure 4. Average daily trends in mobility in Malaysia

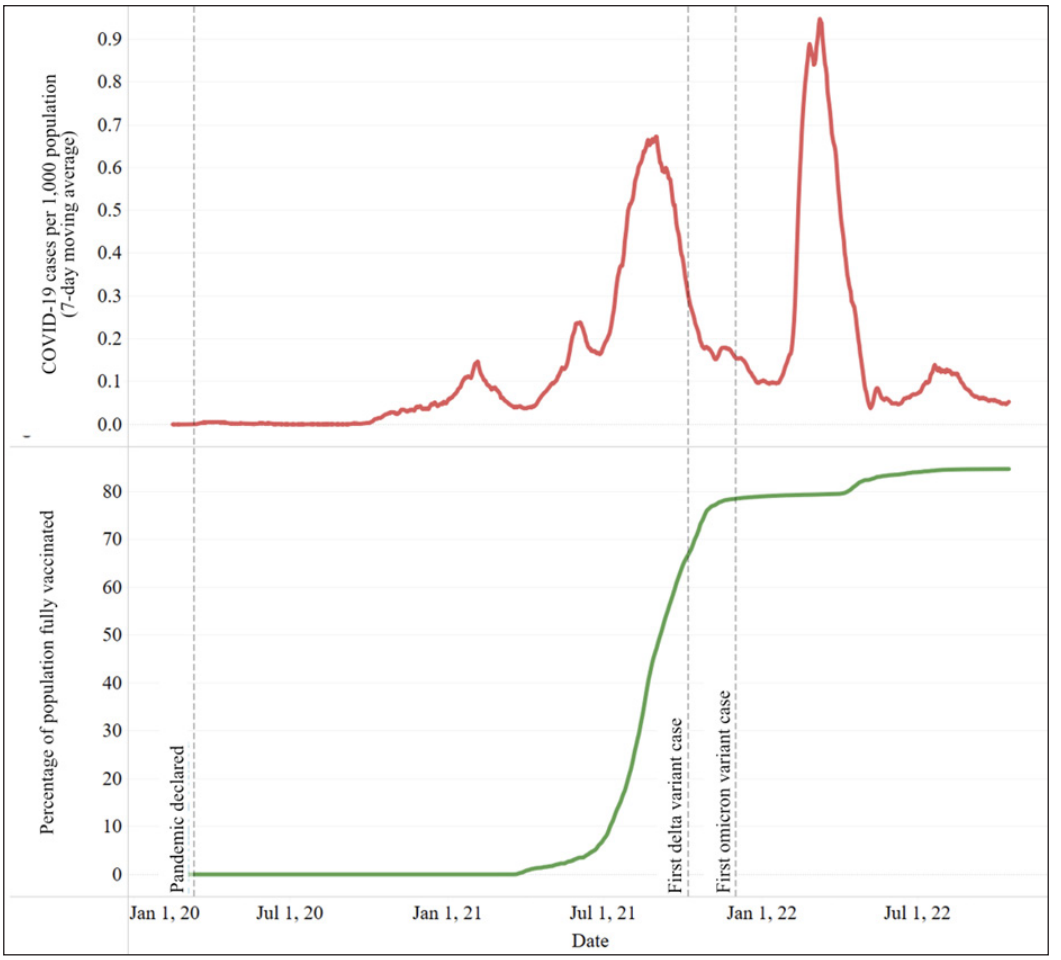


Figure 5. Average daily COVID-19 cases and vaccination trends in Malaysia

STATISTICAL METHODOLOGY

Separate regression models were estimated for each trip type to discern the impacts of different time periods on mobility patterns while controlling for other important variables. The analysis data were arranged based on the state-day level. In each of the models, the dependent variable was the percentage change in the number of trips as compared to the median value of January to early February 2020 data for the same day of the week. These variables were continuous in nature and exhibited normal distribution trends, with linear relationships with the predictor variables. Thus, linear regression models were estimated for each trip category.

Additionally, concerns arise due to repeated observation within each state over time, which may result in both spatial and temporal correlations in the dataset. This tends to

result in specific states showing a higher (or lower) number of trips made compared to other similar states. Failure to account for this correlation would result in a biased coefficient estimate, resulting in incorrect inferences. Thus, a mixed-effects framework was utilized, which allows for the estimation of a series of state-specific random effects, thus allowing the model to account for factors that are not being captured by the model directly due to state-specific variation. The general form of the model is shown in Equation 2:

$$y_{it} = \beta X_{it} + u_i + \varepsilon_{it} \quad [2]$$
$$u_i \sim N(0,\sigma_u^2)$$
$$\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$$

where y_{it} is the percentage change in the mobility for state i on day t; β is the estimated coefficient from the model; X_{it} is the predictor variables; u_i is the random effect that captures the unobserved state-level effect (varies from state to state but remains the same within a state); and ε_{it} is the error term.

Within the context of this study, time (i.e., date) was not treated as a random effect since time period-related variables were already included as predictors (i.e., fixed effects), which accounted for these effects in the models.

RESULTS AND DISCUSSION

Table 2 shows the results for mixed-effects linear regression models for the percentage change in various types of mobility between February 2020 and October 2022. When interpreting the results, a positive estimated coefficient indicates that the mobility in a given state increased as that variable increased. Conversely, a negative estimate coefficient demonstrates a reduction in mobility, while a close to zero coefficient indicates no change. The following sub-sections discuss the results obtained and their implications for each of these four categories of variables.

Demographic and Socioeconomic Factors

Several demographic and socioeconomic factors were assessed, including the percentage of the Malay population (the major race in Malaysia), the percentage of the rural population, sex, poverty rate, unemployment rate, and median income. The model results showed that an increase in the poverty rate or unemployment rate resulted in reduced mobility across various trip types, which is supported by prior studies (Bamney et al., 2022). Generally, these two factors were correlated and, thus, were not included in the model simultaneously. A one percent increase in the population living below the national poverty line resulted in a 1.6% and 1.8% reduction in workplace trips and trips to grocery stores and pharmacies, respectively. Similarly, a 1% increase in the unemployment rate resulted in a 2.4% reduction

in trips to transit stations. Retail and recreational trips did not seem to be impacted by these factors.

Table 2
Results of the mixed-effect linear regression model

Variable	Estimate coefficient			
	Workplace (n = 15,567)	**Transit station (n = 15,465)**	**Retail and recreation (n = 15,515)**	**Grocery and pharmacy (n = 15,508)**
Intercept	-1.622	961.454*	173.993*	9.532*
Sociodemographic factors				
Percentage of population below the national poverty line	-1.553*	N/A	N/A	-1.786*
Percentage of unemployment	N/A	-2.377*	N/A	N/A
Median income in RM1,000	1.771*	N/A	N/A	N/A
Percentage of the rural population	0.757*	N/A	N/A	N/A
Percentage of the Malay population	-0.164*	-1.186*	0.364*	N/A
Percentage of the female population	N/A	-18.405*	-4.114*	N/A
Timeline				
Pre-pandemic (Feb. 15 - Mar. 10, 2020)	Baseline			
No lockdown (Mar. 11 - Mar. 17, 2020)	-6.531*	8.941*	-1.746	10.112*
MCO-1, nationwide lockdown (Mar. 18 - May 03, 2020)	-57.166*	-74.061*	-74.184*	-43.364*
CMCO, varies by state and date	-28.086*	-46.223*	-32.663*	-2.772*
RMCO, varies by state and date	-19.064*	-25.772*	-16.916*	5.041*
MCO-2, varies by state (Jan. 13 – Mar. 4, 2021)	-36.487*	-54.283*	-43.887*	-12.986*
MCO-3, nationwide lockdown (May 12 – Jun. 28, 2021)	-44.260*	-57.843*	-48.764*	-10.276*
NRP-1 (Jun. 29 – Sep. 30, 2021)	-40.801*	-47.431*	-42.453*	-4.160*
NRP-2 (Jul. 05 – Oct. 17, 2021)	-33.615*	-46.344*	-32.818*	1.963*
NRP-3 (Aug. 04 – Dec. 31, 2021)	-26.785*	-18.603*	-9.706*	16.110*
NRP-4 (Aug. 26, 2021 – Mar. 31, 2022)	-17.595*	-3.724*	1.676*	20.816*
Post-pandemic (Apr. 2022)	-7.075*	-7.383*	1.072	31.747*
Post-pandemic (May 2022)	-19.216*	22.968*	13.524*	32.013*
Post-pandemic (Jun. – Oct. 2022)	-7.714*	N/A	N/A	N/A
Post-pandemic (Jun. 2022)	N/A	15.628*	11.305*	24.096*
Post-pandemic (Jul. 2022)	N/A	13.336*	8.987*	23.285*
Post-pandemic (Aug. 2022)	N/A	6.126*	9.448*	22.691*
Post-pandemic (Sep. 2022)	N/A	16.829*	12.025*	21.976*
Post-pandemic (Oct. 2022)	N/A	13.020*	10.730*	21.211*

Table 2 *(continue)*

Variable	Estimate coefficient			
	Workplace (n = 15,567)	**Transit station (n = 15,465)**	**Retail and recreation (n = 15,515)**	**Grocery and pharmacy (n = 15,508)**
COVID-19 information				
New COVID-19 cases per 100,000 population	N/A	-0.096*	-0.068*	-0.020*
New COVID-19 death per 100,000 population	-2.482*	N/A	N/A	N/A
Day of week				
Sunday		Baseline		
Monday	-5.325*	4.125*	2.473*	7.190*
Tuesday	-3.998*	3.207*	1.412*	6.256*
Wednesday	-2.903*	4.258*	2.046*	7.732*
Thursday	-3.069*	2.291*	1.726*	6.426*
Friday	0.003	1.429*	0.506	5.038*
Saturday	3.090*	1.759*	-0.054	1.875*
Variance of intercept, $\sigma^2(u_i)$	63.950	2749.900	99.760	234.700

Note. MCO = Movement Control Order; CMCO = Conditional Movement Control Order; RMCO = Recovery Movement Control Order; NRP = National Recovery Plan; Std. Err. = Standard error; *significant at 95% confidence; N/A = Not applicable due to multicollinearity or insignificant results

States with higher median income tended to have higher mobility related to the workplace, which is aligned with prior research (Beck & Hensher, 2020b). The workplace mobility increased significantly by 1.8% for every RM1,000 increase in median income. One potential factor contributing to this trend in Malaysia is the concentration of major corporations in states with higher income levels (e.g., Selangor, Johor, Pulau Pinang, and the three federal territories) (DOSM, 2023). Similar to prior studies (Bamney et al., 2021, 2022), states with a higher percentage of people living in rural areas tended to experience higher mobility to the workplace when compared to states with higher urban concentration. A 1% increase in the percentage of the rural population would increase mobility by 0.8%. Neither median income nor percentage of rural population showed a significant relationship with non-work trips.

The results also showed that the states with a higher percentage of Malay population tended to have lower mobility related to workplace and transit stations, and more trips for retail and recreational purposes throughout the study period. On average, the mobility for workplace and transit station decreased by 0.16 and 1.19%, respectively, and increased by 0.36% for retail and recreational trips for a 1% increase in the Malay population. For workplace trips, this could be attributed to the increase in unemployment rate among the

Malay demographic, notably among those who graduated in 2020. According to the DOSM, the unemployed graduates increased by 22.5% between 2019 and 2020, with 70% of those comprising Malay graduates (Bernama, 2021).

Sex did not seem to have a significant effect on workplace mobility. However, states with a higher female population tended to have lower mobility to transit stations. This trend contrasts with the observation on-site, where 62% of regular public transport users were female ("Making public transport", 2023). Although it should be noted that the analysed data included the entire female population in Malaysia, whereas the public transport users focused only on certain age groups.

COVID-19 Information and Day of Week

The daily announcements of new COVID-19 cases and deaths influenced mobility across various trip types. According to Table 2, mobility to workplaces decreased significantly by 2.5% with an increase in new COVID-19 deaths for every 100,000 population. For non-work trips, a reduction in mobility was observed as the number of COVID-19 cases increased. On average, for every unit increase in COVID-19 cases per 100,000 population, mobility to transit stations, retail and recreational mobility, and trips to grocery stores and pharmacies decreased by 0.1%, 0.1%, and 0.02%, respectively. These trends are consistent with prior studies (Bamney et al., 2021, 2022). When controlled for the number of COVID-19 cases and deaths, the vaccination rates did not show any significant relationship with any of the trip categories. The primary reason for this trend may be due to the strict lockdown that was still in place during the distribution of the COVID-19 vaccine, as shown in Figure 1.

The results were also controlled for day of the week variables. Results showed that weekdays generally had lower work trips compared to weekends (Saturday and Sunday). However, work trips on Saturday were significantly higher than on Sunday. This trend is somewhat unexpected but may be driven by some states (i.e., Terengganu, Kelantan, Kedah, and Johor) not observing Sunday as a holiday. The other trip categories, on the other hand, exhibited higher mobility on weekdays compared to weekends. This again supports the prior discussion that a reduction in work trips, possibly due to more people working from home or in a hybrid arrangement, increases the non-work trips due to more time available for such activities during weekdays.

Pandemic Period

The timeline was divided into three broad categories: before, during, and after the COVID-19 pandemic. To interpret the timeline variables, each variable is compared to the baseline period, which in this case is the mobility data before the WHO announced the COVID-19 outbreak as a global pandemic.

The model results showed that the pre-pandemic period exhibited approximately 1.6% (i.e., based on the intercept value) reduction in work trips when compared to the January/ February data. However, this result was not statistically significant. Meanwhile, mobility during non-work trips was significantly higher in the same time periods.

After COVID-19 was declared a pandemic, mobility to the workplace saw a reduction of 6.5% before the first lockdown took place in Malaysia (March 11 to March 17, 2020). However, trips to transit stations, grocery stores, and pharmacies showed an increase. Retail and recreational trips did not show any significant change during this period. The fear of the COVID-19 pandemic during the initial stages explains a reduction in work, retail and recreational trips, and an increase in trips to grocery stores and pharmacies to stock up on supplies for the upcoming lockdowns.

As the government announced the first MCO, Malaysia saw the most significant reduction in mobility across all trip types. Work trips reduced by 57% which was directly a result of various economic sectors being forced to shut down, forcing people to WFH. Trips to transit stations, retail and recreational trips reduced by 74% each, whereas trips to grocery stores and pharmacies showed slightly lower reductions at 43% compared to other trip categories. One reason for these trends, particularly for transit station trips, was the closure of all government and private sector premises ("14-day Movement Control Order", 2020), which are primarily concentrated within the vicinity of most public transit systems. Additionally, this may be due to the limited operating time of public transit during the lockdown period ("New schedule", 2020), resulting in a higher reduction in transit-related mobility. Whereas the closure of most services within this sector, except for restaurants, for a take-away option only, might explain the trends observed for retail and recreational trips.

As the number of COVID-19 cases decreased, the government began to ease some restrictions, and Malaysia transitioned into the CMCO phase, followed by the RMCO phase. During these two phases, the mobility across all trip types started to increase (i.e., the estimated coefficients are less negative). During the CMCO phase, the mobility to the workplace increased by 29.1%, and during the RMCO, it increased by another 9.0%. This increase in mobility to the workplace during these two phases may be driven by economic sectors, most of which were allowed to operate after the first MCO. Similarly, transit station trips, retail and recreational trips, and trips to grocery stores and pharmacies increased by 27.8, 41.5, and 40.6%, respectively, during the CMCO phase. During the RMCO phase, trips across these three categories increased further by 20.5, 15.7, and 7.8%, respectively. However, it should be noted that during both CMCO and RMCO phases, the mobility across all trip types was still lower than the mobility during the baseline period, except for trips to grocery stores and pharmacies. Grocery and pharmacy-related trips were consistently higher than the other trip categories. Moreover, the number of these trips was actually higher by 5% during the RMCO phase compared to the pre-pandemic period. These trends

are expected, as these trips were considered essential trips and people were expected to make more such trips during relaxed lockdown periods.

It should be noted that both CMCO and RMCO phases varied by state and date, and the implementation of these phases depended on the number of COVID-19 cases recorded. As the number of COVID-19 cases started to increase again, a second MCO (MCO-2) was reintroduced in all localities, except for the state of Sarawak, on January 13, 2021. As expected, mobility decreased significantly across all trip purposes during this period. Due to the surge in COVID-19 cases, the third MCO was reinstated, resulting in further reduction in mobility. However, during both MCO-2 and MCO-3, the reductions in trips were less pronounced compared to MCO-1. This was primarily due to the level of restrictions imposed and the nature of the implementation of these lockdowns. For example, all government and private sector premises were not allowed to operate during the first MCO, except for essential services ("14-day Movement Control Order", 2020). However, the second MCO saw some relaxation given to key economic sectors to operate with strict SOP ("Only 5 economic sectors", 2021). Moreover, unlike the first MCO, the second MCO had a different start date for each state.

As the COVID-19 pandemic started to get under control, the nation implemented the NRP in four phases. In terms of restrictions, NRP-1 had similar restrictions to MCO. However, more industrial, business, and commercial sectors were allowed to operate compared to the MCO phase. Thus, the NRP phases showed consistent recovery in mobility across all trip types. For example, between NRP-1 and NRP-2, the workplace-related mobility increased significantly by 7.2%; further increased by 6.8% between NRP-2 and NRP-3; and increased by 9.2% between NRP-3 and NRP-4. However, the workplace mobility during these transitions was still lower than the baseline. Similar trends were also observed for transit trips and retail and recreational trips. Trips to grocery and pharmacies, on the other hand, showed a much rapid increase and quickly exceeded the trips during the pre-pandemic period.

It is worth noting that various types of restriction variables (e.g., school closures, interstate travel bans, limited travel radius, and restaurant closures) were also analysed as a part of this study. However, since the time period variables accounted for these effects, correlations between independent variables were observed when types of restrictions were included in the models. Hence, these factors were not considered in the final model.

Post-Pandemic Period

The post-pandemic period was analysed by month to see if there were any changes in mobility as time progressed. For the workplace mobility model, three different categorical variables were created for this period: April 2022, May 2022, and June to October 2022. After the government announced the transition to the endemic phase on April 1, the mobility

to the workplace increased significantly by 10.5% in the month of April 2022 compared to the NRP-4 phase. However, in May 2022, the mobility reduced again by 12.1%, and this could be mainly due to the national public holiday observed during this month. Between June and October 2022, mobility increased by 11.5% compared to May 2022. Interestingly, other trip categories, including transit station trips during the same time period, showed a marked increase. Beginning in April 2022, the transit station mobility began to increase, with the period between May and October 2022 showing even higher mobility than the pre-pandemic periods. Similar trends were also observed for retail and recreational trips, as well as trips to grocery stores and pharmacies, the latter of which showed the greatest increase in mobility during the post-pandemic period.

It should be noted that all restrictions that were placed during the pandemic period were lifted on April 1, 2022. Despite this fact, work trips showed a much lower recovery even after the pandemic. On the other hand, non-work trips exceeded pre-pandemic levels between April and October 2022. Since the model captures the effects of unemployment on mobility explicitly or implicitly (through other variables such as poverty rate, median income), it is reasonable to assume that these changes in mobility are largely due to WFH activity. Lower levels of work trips, even after the pandemic, are thus likely due to increased WFH or hybrid work arrangements. As employees choose to WFH, they are more likely to engage in additional non-work trips, as the time saved from commuting can be reallocated to other activities (Moreno & Moeckel, 2016), which explains greater recovery among other trip categories. Prior studies also support this assertion. A local study utilizing a survey-based approach found that the WFH policy would not negatively impact the employees' efficacy (Prihadi et al., 2021). Another survey-based study conducted by the Malaysian Employers Federation found that about 62% of Malaysian companies that participated in the survey preferred a hybrid work arrangement (Malaysian Employers Federation [MEF], 2021). This might explain why more companies prefer their employees to WFH or have a hybrid work arrangement even after the pandemic.

From a transportation engineering and planning perspective, the persistent reduction in work-related trips has several implications. First, this could impact the future of the transportation system and land use in Malaysia. The shift toward remote and hybrid work arrangements may reduce peak-hour commuting demand, alter traffic patterns, and potentially decrease the need for large-scale office districts. Second, the four-step traditional demand forecasting model may need to be recalibrated to account for individuals working from home. This is particularly true for the trip generation model, where the probability of an individual working from home needs to be estimated (Moeckel, 2017). Additionally, this could also impact the mode choice model based on the increasing trend of mobility to transit stations, as also shown in prior studies (Hensher et al., 2022; Loa & Habib, 2023). Third, from a transportation infrastructure investment perspective, a continued decline

in work trips could lead to financial risks if demand projections fail to account for these shifts. Historical evidence has shown that many road and rail projects tend to overestimate future travel demand (Flyvbjerg et al., 2006). With more employees working remotely, such overestimations could become even more pronounced, leading to underutilized infrastructure and misallocated resources. Furthermore, beyond transportation, these trends have broader implications for office-based employment. Given the pace at which technology is advancing, flexible work arrangements may be encouraged by both employers and employees, which could reduce demand for traditional office space. Cities must adapt by ensuring that commercial real estate policies, zoning regulations, and transit planning align with these evolving work patterns.

CONCLUSION

The spread of the COVID-19 pandemic introduced unprecedented changes globally in the transportation sector. This study provides important insights into the impacts of the COVID-19 pandemic on mobility trends in Malaysia. The mobility trend data were obtained from Google's community mobility reports for the period from February 2020 to October 2022 at 24-hour aggregation for each state. Four different types of trips were investigated: work trips, trips to transit stations, retail and recreational trips, and trips to grocery stores and pharmacies. Additional data related to the number of COVID cases, vaccines administered, and COVID-related deaths were obtained for every state for each day. Census data for Malaysia were also obtained and integrated to get demographic details for each state. Lastly, details of lockdowns introduced by the government during the study period were also obtained.

The investigations showed interesting trends in mobility throughout the pandemic and the early stages of the post-pandemic. During March 2020, when COVID-19 started to spread and the WHO declared it a pandemic, the government introduced a series of MCOs, which resulted in a sharp decline in all types of trips. Work trips were reduced by 57%, while trips to transit stations and retail and recreational trips were reduced by 74%. Essential trips, including trips to grocery stores and pharmacies, were reduced by the least amount at 43%. With time, the restrictions were made less strict, resulting in an increase in mobility and the number of COVID-19 cases across the country. Similarly, in 2021, the changes in mobility across all categories were found to be correlated with the severity of lockdown and the number of COVID-19 cases, and population demographics. After the transition to the endemic phase, an increase in mobility was observed for non-work trips, but a reduction was observed for workplace trips.

One key finding in this study is the differential recovery rates of work and non-work trips following the post-pandemic period, highlighting the lasting impact of the COVID-19 pandemic on travel behaviour. While both types of trips are beginning to rebound, work

trips are recovering at a slower rate than non-work trips. These results suggest that changes in work arrangements, such as hybrid and remote work models, may continue to shape urban mobility patterns in the long run. The implications of this trend are significant and provide insight for traffic engineers and city planners to consider when weighing the travel needs and expected traffic generation associated with the new travel patterns in Malaysia. For example, urban centres may experience prolonged periods of reduced congestion, which could influence future transportation planning and infrastructure investments. Public transportation systems may also need to adapt to changes in demand and peak usage times. A related change could be in commercial and real estate demand, as demand for commercial spaces could decrease. Overall, the pandemic resulted in a major shift in transportation and mobility, and its effect may last well into the future. Understanding these shifts in mobility patterns can help policymakers design more resilient transportation systems to mitigate the impact of future crises, whether health-related or otherwise. Future research can build on these findings to explore the permanence of these behavioural shifts and their implications for urban planning, public transportation, and workplace policies.

While this study provides crucial insights into the changes in travel patterns due to the COVID-19 pandemic, there are several important limitations that are worth noting. First, the analysis uses data drawn from Google Mobility Reports, which might represent a subset of trips. Moreover, the data primarily comes from smartphone users with Google Location History enabled, which may lead to an underrepresentation of certain population demographics, such as low-income populations or people concerned with data privacy. Nevertheless, the study provides important insights into the overall mobility trends in Malaysia during and after the pandemic. Future studies could supplement this data with surveys or official transportation records to enhance representativeness. Another limitation was the selection of January and February 2020 as the baseline period, due to data constraints, despite the widespread news of COVID-19 spreading globally in late 2019. Although it is important to consider the varying impact of different time periods, this timeframe still represents a relevant reference point just before the country experienced significant impacts of COVID-19 on mobility. For future study, a longer baseline period should be considered, considering the availability of data. Additionally, a concern arises about the potential overlap between trips to workplaces and transit stations (i.e., a work trip might also be a trip to a transit station). However, a consistent trend was shown for other non-work trips, where significant increases in mobility were observed after the pandemic. Lastly, data on the percentage of the population working from home in Malaysia during various time periods were not available. This makes it challenging to determine the exact reasons for the reduction in work trips during various lockdown phases. However, since the analysis in this study controls for the impacts of unemployment on mobility, it is reasonable to assume that such reductions are largely due to increases in WFH activity.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the Department of Statistics Malaysia for providing the demographic data and the Ministry of Health Malaysia for supplying the COVID-19 case data. The interpretations and conclusions in this study are solely those of the authors and do not necessarily reflect the views of these organisations.

REFERENCES

14-day Movement Control Order begins nationwide on Wednesday. (2020, March 16). *New Straits Times*. https://www.nst.com.my/news/nation/2020/03/575180/14-day-movement-control-order-begins-nationwide-wednesday

Anand, R. (2021, February 24). PM Muhyiddin receives first COVID-19 vaccine as Malaysia kicks off mass inoculation campaign. *The Straits Times*. https://www.straitstimes.com/asia/se-asia/pm-muhyiddin-receives-first-covid-19-vaccine-as-malaysia-kicks-off-mass-inoculation

Awad-Núñez, S., Julio, R., Gomez, J., Moya-Gómez, B., & González, J. S. (2021). Post-COVID-19 travel behaviour patterns: Impact on the willingness to pay of users of public transport and shared mobility services in Spain. *European Transport Research Review*, *13*, 20. https://doi.org/10.1186/s12544-021-00476-4

Badr, H. S., Du, H., Marshall, M., Dong, E., Squire, M. M., & Gardner, L. M. (2020). Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study. *The Lancet Infectious Diseases, 20*(11), 1247–1254. https://doi.org/10.1016/S1473-3099(20)30553-3

Balbontin, C., Hensher, D. A., & Beck, M. J. (2024). Relationship between commuting and non-commuting travel activity under the growing incidence of working from home and people's attitudes towards COVID-19. *Transportation*, *51*, 2225–2251. https://doi.org/10.1007/s11116-023-10403-2

Bamney, A., Gupta, N. Jashami, H., Megat-Johari, M. U., & Savolainen. P. (2022). An analysis of changes in county-level travel behavior considering COVID-19–related travel restrictions, immunization patterns, and political leanings. *Journal of Transportation Engineering, Part A: Systems*, *148*(11), 04022096. https://doi.org/10.1061/JTEPBS.0000748

Bamney, A., Jashami, H., Pantangi, S. S., Ambabo, J., Megat-Johari, M. U., Cai, Q., Gupta, N., & Savolainen, P. T. (2021). Examining impacts of COVID-19-related stay-at-home orders through a two-way random effects model. *Transportation Research Record*, *2677*(4), 255–266. https://doi.org/10.1177/03611981211046921

Beck, M. J., & Hensher, D. A. (2020a). Insights into the impact of COVID-19 on household travel and activities in Australia – The early days of easing restrictions. *Transport Policy*, *99*, 95–119. https://doi.org/10.1016/j.tranpol.2020.08.004

Beck, M. J., & Hensher, D. A. (2020b). Insights into the impact of COVID-19 on household travel and activities in Australia – The early days under restrictions. *Transport Policy*, *96*, 76–93. https://doi.org/10.1016/j.tranpol.2020.07.001

Bergman, N. K., & Fishman, R. (2023). Correlations of mobility and COVID-19 transmission in global data. *PLOS One*, *18*(7), e0279484. https://doi.org/10.1371/journal.pone.0279484

Bernama. (2021, September 23). Teraju: Pandemic causes drop in hiring of Bumiputera grads in 2020. *New Straits Times*. https://www.nst.com.my/news/nation/2021/09/729975/teraju-pandemic-causes-drop-hiring-bumiputera-grads-2020

Bonera, M., & Martinelli, V. (2023). COVID-19 and public transport: Two years later. Investigating the transport demand trend in the city of Brescia. *Transportation Research Procedia*, *69*, 376–383. https://doi.org/10.1016/j.trpro.2023.02.185

Chang, S., Pierson, E., Koh, P.W., Gerardin, J., Redbird, B., Grusky, D., & Leskovec, J. (2021). Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, *589*, 82–87. https://doi.org/10.1038/s41586-020-2923-3

Currie, G., Jain, T., & Aston, L. (2021). Evidence of a post-COVID change in travel behaviour – Self-reported expectations of commuting in Melbourne. *Transportation Research Part A: Policy and Practice*, *153*, 218–234. https://doi.org/10.1016/j.tra.2021.09.009

Department of Statistics Malaysia. (2021). *Household income estimates and incidence of poverty report Malaysia, 2020*. DOSM. http://statistics.gov.my/site/downloadrelease?id=household-income-estimates-and-incidence-of-poverty-report-malaysia-2020&lang=English&admin_view=

Department of Statistics Malaysia. (2023). *Household income and expenditure*. DOSM. https://open.dosm.gov.my/dashboard/household-income-expenditure

Department of Statistics Malaysia. (2025). *Labour markets*. DOSM. https://open.dosm.gov.my/dashboard/labour-market

Department for Transport. (2024). *Official statistics domestic transport usage by mode*. DfT. https://www.gov.uk/government/statistics/transport-use-during-the-coronavirus-covid-19-pandemic/domestic-transport-usage-by-mode

Dias, C., Abdul Rahman, N., Abdullah, M., & Sukor, N. S. A. (2021). Influence of COVID-19 mobility-restricting policies on individual travel behavior in Malaysia. *Sustainability*, *13*(24), 13960. https://doi.org/10.3390/su132413960

Flyvbjerg, B., Holm, M. K. S., & Buhl, S. L. (2006). Inaccuracy in traffic forecasts. *Transport Reviews*, *26*(1), 1–24. https://doi.org/10.1080/01441640500124779

Google. (2022). *COVID-19 community mobility reports*. https://www.google.com/covid19/mobility/

Gramsch, B., Guevara, C. A., Munizaga, M., Schwartz, D., & Tirachini, A. (2022). The effect of dynamic lockdowns on public transport demand in times of COVID-19: Evidence from smartcard data. *Transport Policy*, *126*, 136–150. https://doi.org/10.1016/j.tranpol.2022.06.012

Haseeb, A., & Mitra, R. (2024). Travel behaviour changes among post-secondary students after COVID-19 pandemic – A case of Greater Toronto and Hamilton area, Canada. *Case Studies on Transport Policy*, *17*, 101245. https://doi.org/10.1016/j.cstp.2024.101245

Hensher, D. A., Balbontin, C., Beck, M. J., & Wei, E. (2022). The impact of working from home on modal commuting choice response during COVID-19: Implications for two metropolitan areas in Australia. *Transportation Research Part A: Policy and Practice*, *155*, 179–201. https://doi.org/10.1016/j.tra.2021.11.011

Iglesias, V., & Raveau, S. (2024). Effect of the COVID-19 pandemic on crowding aversion in public transport and transport mode choice: The case of Santiago, Chile. *Transport Policy*, *146*, 167–174. https://doi.org/10.1016/j.tranpol.2023.11.017

International Organization of Migration. (n.d.). *DTM (COVID-19) global mobility restrictions overview: March 2020 - January 2023*. IOM. https://dtm.iom.int/reports/dtm-covid-19-global-mobility-restrictions-overview-march-2020-january-2023

Kar, A., Le, H. T. K., & Miller, H. J. (2022). What is essential travel? Socioeconomic differences in travel demand in Columbus, Ohio, during the COVID-19 lockdown. *Annals of the American Association of Geographers, 112*(4), 1023–1046. https://doi.org/10.1080/24694452.2021.1956876

Loa, P., & Habib, K. N. (2023). Identifying the determinants of anticipated post-pandemic mode choices in the Greater Toronto area: A stated preference study. *Transportation Research Record*, *2677*(6), 199–217. https://doi.org/10.1177/03611981221145133

Lodi, C., Marin, G., Polidori, P., & Teobaldelli, D. (2024). Students' commuting habits to the university: Transportation choices during the COVID-19 era. *Case Studies on Transport Policy*, *17*, 101217. https://doi.org/10.1016/j.cstp.2024.101217

Making public transport facilities safer for all. (2023, November 15). *The Stars*. https://www.thestar.com.my/news/nation/2023/11/15/making-public-transport-facilities-safer-for-all#:~:text=Loke%20also%20said%20that%2062,21%20and%2030%20were%20women

Malaysian Employers Federation. (2021). *Survey on implementation of work from home and work from office practices in response to COVID-19 pandemic*. MEF. https://mefacademyold.mef.org.my/publications/publication_info.aspx?ID=65

Matson, G., McElroy, S., Circella, G., & Lee, Y. (2021). *Telecommuting rates during the pandemic differ by job type, income, and gender*. National Center for Sustainable Transportation. https://doi.org/10.7922/G2445JT6

Ministry of Health Malaysia. (n.d.). *Open data on COVID-19 in Malaysia*. MOH. https://github.com/MoH-Malaysia/covid19-public/

Moeckel, R. (2017). Working from home: Modeling the impact of telework on transportation and land use. *Transportation Research Procedia*, *26*, 207–214. https://doi.org/10.1016/j.trpro.2017.07.021

Moreno, A. T., & Moeckel, R. (2016, July 10-15). *Microscopic destination choice: Incorporating travel time budgets as constraints* [Paper presentation]. In World Conference on Transport Research, Shanghai, China. https://mediatum.ub.tum.de/doc/1430699/403854535069.pdf

Mukhtar, M., Lekshmipathy, J., & Suthar, M. (2024). Analyzing the impact of COVID-19 pandemic on mobility behavior and formulation of post-COVID public transport strategy. In *AIP Conference Proceedings* (Vol. 3050, No. 1, p. 030001). AIP Publishing. https://doi.org/10.1063/5.0193896

New schedule for Prasarana services during MCO. (2020, March 23). *New Straits Times.* https://www.nst.com.my/news/nation/2020/03/577378/new-schedule-prasarana-services-during-mco

Ong, F., Loa, P., & Habib, K. N. (2024). A behavioural analysis of post-pandemic modality profiles for non-commuting trips in the Greater Toronto Area. *Travel Behaviour and Society*, *34*, 100690. https://doi.org/10.1016/j.tbs.2023.100690

Only 5 economic sectors allowed to operate during MCO - PM. (2021, November 1). *The Sun*. https://thesun.my/home_news/only-5-economic-sectors-allowed-to-operate-during-mco-pm-FL6049337

Pawar, D. S., Yadav, A. K., Choudhary, P., & Velaga, N. R. (2021). Modelling work-and non-work-based trip patterns during transition to lockdown period of COVID-19 pandemic in India. *Travel Behaviour and Society*, *24*, 46–56. https://doi.org/10.1016/j.tbs.2021.02.002

Politis, I., Georgiadis, G., Nikolaidou, A., Kopsacheilis, A., Fyrogenis, I., Sdoukopoulos, A., Verani, E., & Papadopoulos, E. (2021). Mapping travel behavior changes during the COVID-19 lockdown: A socioeconomic analysis in Greece. *European Transport Research Review*, *13*, 21. https://doi.org/10.1186/s12544-021-00481-7

Prihadi, K. D., Lim, E. S. Z., Chan, K. C., Lee, S. M. H., & Ridwan, A. (2021). Efficacy of working from home among urban professionals in Malaysia during the pandemic: The robust predictive role of mattering. *International Journal of Public Health Science*, *10*(1), 215–220. https://doi.org/10.11591/ijphs.v10i1.20736

Przybylowski, A., Stelmak, S., & Suchanek, M. (2021). Mobility behaviour in view of the impact of the COVID-19 pandemic — Public transport users in Gdansk case study. *Sustainability*, *13*(1), 364. https://doi.org/10.3390/su13010364

Sharma, S. V., Chuang, R.-J, Rushing, M., Naylor, B., Ranjit, N., Pomeroy, M., & Markham, C. (2020). Social determinants of health–related needs during COVID-19 among low-income households with children. *Preventing Chronic Diseases*, *17*, 200322. https://doi.org/10.5888/pcd17.200322

Shibayama, T., Sandholzer, F., Laa, B., & Brezina, T. (2021). Impact of COVID-19 lockdown on commuting: A multi-country perspective. *European Journal of Transport and Infrastructure Research*, *21*(1), 70–93. https://doi.org/10.18757/ejtir.2021.21.1.5135

Soltani, A., Azmoodeh, M., Doostvandi, M., Ahmadi, A. S., & Rahimi, M. (2024). Post-COVID-19 campus commuting patterns and influential factors: Evidence from a developing country. *Transportation Planning and Technology*, *47*(4), 566–597. https://doi.org/10.1080/03081060.2023.2300800

Tan, S., Fang, K., & Lester, T. W. (2023). Post-pandemic travel patterns of remote tech workers. *Transportation Research Interdisciplinary Perspectives*, *19*, 100804. https://doi.org/10.1016/j.trip.2023.100804

Tang, A. (2020, March 16). Malaysia announces Movement Control Order after spike in COVID-19 cases (updated). *The Star*. https://www.thestar.com.my/news/nation/2020/03/16/malaysia-announces-restricted-movement-measure-after-spike-in-covid-19-cases

World Health Organization. (2020). *WHO Director-General's opening remarks at the media briefing on COVID-19*. WHO. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

Xi, H., Li, Q., Hensher, D. A., Nelson, J. D., & Ho, C. (2023). Quantifying the impact of COVID-19 on travel behavior in different socio-economic segments. *Transport Policy*, *136*, 98–112. https://doi.org/10.1016/j.tranpol.2023.03.014

Zhang, M., Wang, S., Hu, T., Fu, X., Wang, X., Hu, Y., Halloran, B., Li, Z., Cui, Y., Liu, H., Liu, Z., & Bao, S. (2022). Human mobility and COVID-19 transmission: A systematic review and future directions. *Annals of GIS*, *28*(4), 501–514. https://doi.org/10.1080/19475683.2022.2041725

Pertanika J. Sci. & Technol. 33 (6): 2641 - 2663 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Review Article

Credit Card Fraud Detection Using Machine Learning and Deep Learning Techniques: A Review

Shuchita Sheokand and Sunita Beniwal*

Department of Computer Science and Engineering, Guru Jambheshwar University of Science and Technology, Hisar, Haryana 125001, India

ABSTRACT

Credit cards are becoming increasingly common worldwide due to the growing demand for online payments and shopping, resulting in a rise in credit card usage. Consequently, several opportunities may arise for fraudsters to cause scams. A credit card fraud detection (CCFD) system needs to be established, which alerts banking organizations when fraudulent activities occur. Many machine learning (ML) and deep learning (DL) models have been proposed by researchers to detect credit card fraud. In this survey, work done on CCFD has been reviewed. Analyzing the papers provides insight into the challenges associated with CCFD, such as class imbalance and the performance of various ML and DL techniques in detecting fraud. The techniques most used are clustering, DL, ensemble, ML, optimization, synthetic minority oversampling technique (SMOTE), and transfer learning. The analysis was done on datasets, evaluation metrics, tools, publication year, and techniques commonly used for CCFD. The papers studied have utilized Python, MATLAB, and Apache on datasets for credit card fraud taken from various sources, including Kaggle, e-commerce transactions, insurance fraud, and real-world credit card transaction datasets. A comparison between various techniques is conducted using different metrics, including accuracy, precision, recall, and *F*-measure.

Keywords: Clustering, credit card fraud detection, deep learning, machine learning, optimization

ARTICLE INFO

Article history:
Received: 09 December 2024
Accepted: 29 May 2025
Published: 08 October 2025

DOI: https://doi.org/10.47836/pjst.33.6.11

E-mail addresses:
shuchita.sheokand@gmail.com (Shuchita Sheokand)
sunitabeniwalcse@gmail.com (Sunita Beniwal)
* Corresponding author

INTRODUCTION

CCFD is a significant concern in the financial industry, garnering the attention of ML and statistical intelligence groups (Afriyie et al., 2023; Mienye & Sun, 2023), which has resulted in numerous methods being proposed to address this issue (Ahmad et al., 2023; Mahmoudi & Duman, 2015). Given

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

the problem of class imbalance, where the number of unaffected transactions greatly exceeds the number of scams, and concept drift (Alhashmi et al., 2023), where transactions may alter their numerical assets over time, this really seems to be most challenging from a learning viewpoint. Although these are not the only challenges fraud detection systems face when approaching learning (Bahnsen et al., 2015; Jain et al., 2019). Frauds in credit cards can be carried out offline as well as online. E-commerce organizations extensively conduct data mining on their users' logs. This type of log files contains both authentic and fraudulent transactions (Benchaji et al., 2021; Kim et al., 2019). Credit card fraud arises when a fraudster uses an extra card without the approval of a credit card user by locating crucial information like personal identification number (PIN) and password (Carcillo, Le Borgne, et al., 2018; Carcillo et al., 2021; Karthik et al., 2022). As the world is moving towards a cashless society, there is an increasing dependence on online transactions. Modern fraud does not need criminals to be present physically during the fraud, and they employ many approaches to hide their identity (Dal Pozzolo, Caelen, et al., 2014; Dal Pozzolo, Johnson, et al., 2014; Langevin et al., 2022).

Multi-factor authentication (MFA) methods and advanced encryption techniques, such as biometric tools, have been utilized to prevent fraudulent actions and protect credit card holders and issuers, thereby reducing financial losses for users. Conversely, impostors persistently seek out a vulnerability to exploit (Darwish, 2020; Prabhakaran & Nedunchelian, 2023). There is a problem of class imbalance, i.e., distribution among the classes is not even, and the performance of ML algorithms is affected because of this. Various approaches and procedures have been implemented to reduce the disparity between the two classes, addressing the problem. One of the methods used is oversampling to raise the number of transactions in the minority class by arbitrarily imitating the illustrations (Fiore et al., 2019), whereas under-sampling is a method that is utilized to decrease the number of transactions in the majority class (Forough & Momtazi, 2022; Padhi et al., 2022). Ebiaredoh-Mienye et al. (2020, 2022) used a cost-sensitive adaptive boosting (AdaBoost) weighting technique to solve class imbalance by assigning higher weights to the minority class than the majority class, thus increasing accuracy in chronic kidney disease prediction. Obaido et al. (2022) stated that accuracy can be misleading for imbalanced data while using it for disease prediction, so they used balanced accuracy for imbalanced data, which is the arithmetic mean of specificity and sensitivity.

Fraud detection and prevention are the primary methods for reducing credit card scams. To avoid scams, a set of processes, norms, and guidelines is in place. Protecting payment access, intrusion detection schemes, and firewalls are mostly employed approaches to prevent scams (Gama et al., 2014; Zhu et al., 2020). Approaches like data mining, predictive analytics, and clustering methods can be employed for anomaly detection, which you have already employed for disease detection.

Moreover, none of these methods can be used without ML, whether it is unsupervised or supervised, which is valuable for organizing credit card fraud. However, to detect all fraudulent activities, these ML systems must overcome numerous challenges (Ghosh et al., 2022; Habibpour et al., 2023). The primary objective of this study is to analyze different techniques employed in the detection of credit card fraud. The present approaches can be classified into many categories, such as clustering, ML, optimization, SMOTE, transfer learning (TL), ensemble, and DL. The analysis is also conducted by considering various tools, evaluation metrics, and datasets used (Halvaiee & Akbari, 2014).

RELATED WORK

This section provides an analysis of different methods used for detecting credit card fraud. CCFD methods encompass various methodologies, including clustering, ML, optimization, SMOTE, TL, ensemble methods, and DL, among others.

Categorization of Various CCFD Techniques

The categorization of various CCFD methods is: clustering, ML, optimization, SMOTE, TL, ensemble, and DL. Figure 1 shows the categorization of various CCFD methods. The application of these techniques for CCFD and their performance are explained in the ongoing section.

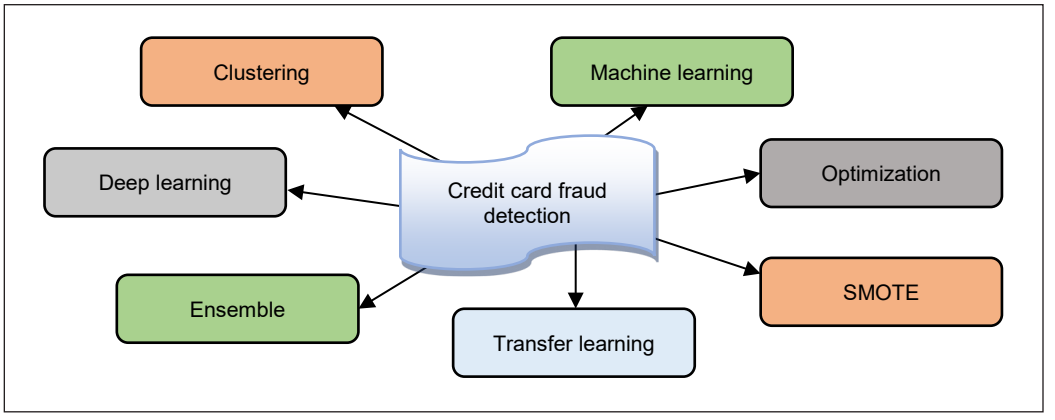


Figure 1. Classification of credit card fraud detection techniques
Note. SMOTE = Synthetic minority oversampling technique

Clustering-Based Methods

Carneiro et al. (2022) introduced a value clustering for categorical attributes (VCCA) method in which the cost is decreased by reducing the training period and by permitting the presence of high-cardinality (HC) features. In this method, ML classifiers like

Support Vector Machines (SVMs) and Random Forests are explored. This approach delivered the programmed design, which was further refined and revised by fraud agents in a transformation board. Kültür and Çağlayan (2017) developed a Cardholder Behavior Model (CBM) and concentrated on real-world issues of CCFD by suggesting an innovative method. The CBM introduced in this paper has a real-world influence and can suggest an extra fraud detection implementation that will work together with the existing rule-based capabilities.

Ahmad et al. (2023) devised an approach of Similarity-based Selection (SBS) to resolve the issue of an imbalanced class by utilizing Fuzzy C-means. To highlight the effectiveness of SBS methodology and validate its advantage, research performance was related to further methods and to address the issue of random under-sampling (RUS), SBS confirms the resemblance and honesty of the features. The determination of discovering the finest under-sampling method is free from RUS issues and gives a better outcome.

DL-Based Methods

The text2IMG conversion method developed by Alharbi et al. (2022) was designed to provide effective outcomes for CCFD. By using computer vision, this technique enhanced the detection of credit card fraud. This method also addresses future approaches for converting various types of text data into images and organizing the information in distinct forms. To discover new differentiating characteristics for CCFD, several feature engineering techniques were used, and this model also offers new features to enhance the performance. Forough and Momtazi (2022) introduced deep neural networks (DNNs) and probabilistic graphical models (PGMs) for analyzing credit fraud detection. This technique was analyzed with the baseline method by utilizing two practical databases and then observed how taking the concealed consecutive relations among businesses and anticipated labels may enhance the outcomes, and this model accomplished the capable outcomes in under-sampling and oversampling.

Li et al. (2020) developed a deep representation learning method for detecting credit card scams, which offered the benefit of achieving better and consistent performance. This method comprises the particularly enhanced loss function, FCL, and DNN. To expand the intra-class density and interclass departure, full center loss (FCL) was capable of managing the deep representation learning technique from a selection of approaches. Competitive swarm optimization-Deep convolutional neural network (CSO-DCNN) was devised by Karthikeyan et al. (2023) to make a DCNN for identifying the abnormalities by using the proposed CSO. In case of fraud situations, supervised learning on previous records was unable to detect a specific accent. They suggested CSO-DCNN, known as an unsupervised learning technique, and customized ReLu by connecting the contributions and productions.

Ensemble-Based Methods

Karthik et al. (2022) established a Hybrid ensemble model of boosting and bagging group classifiers, and the model also used the prominent features of both approaches. Since the issue of data imbalance was addressed by a hybrid method, experiments using Brazilian bank data and University of California, San Diego-Fair Isaac Corporation (UCSD-FICO) data displayed durability. Xie et al. (2021) introduced a heterogeneous ensemble learning model based on data distribution (HELMDD) to resolve the issue of extremely imbalanced information circulation in CCFD. In this method, the resampling method based on majority class data distribution (RMDD) was developed in terms of the majority class, where it was classified into ordinary samples and boundary samples. It produced numerous balanced subgroups in terms of clustering, in which many base classifiers were trained. Thus, this method achieved a high recall rate for the majority class.

The champion-challenger framework was formulated by Kim et al. (2019) for the advancement procedure. It has significant class and rate imbalances, thus necessitating the use of accurate calculation measures. Additionally, these models were utilized in a practical fraud detection system to assess their effectiveness over a one-month period. Alhashmi et al. (2023) used a bank account fraud (BAF) dataset on which an ensemble method was employed, and performance was compared with other classifiers, namely Voting Classifier, Random Forest, Gradient Boosting, Logistic Regression, and Neural Networks, on different metrics such as accuracy, precision, recall, and F1-score. The suggested ensemble model performed better than other classifiers with an accuracy of 98%. The importance of precision-recall trade-offs in fraud detection was also showcased.

ML-Based Approaches

The learning algorithm was developed by Dal Pozzolo et al. (2017), which was essential for providing opinions throughout the learning problem and accepting precise statements. The suggested learning technique involves individually training a classifier on feedback and a classifier on delayed supervised samples, then relating to detecting warnings, which obviously places major importance on feedback. Afriyie et al. (2023) employed Logistic Regression, Decision Tree, and Random Forest approaches to categorize online credit card transactions as either fraudulent or legitimate. The dataset was balanced before creating the models using the under-sampling strategy to guarantee that the model did not only favor the majority class and prevent over-fitting the techniques to the data.

Alfaiz and Fati (2022) demonstrated a method called AllKNN-CatBoost, which pays on stratified K-fold cross-validation together with a real-life dataset of European cardholders for detecting credit card fraud. To find fraudulent transactions, the ML approaches were taken to the test in the first stage, and then they were chosen to be applied once more in the second stage. AdaBoost and Light Gradient Boosting Machine (LightGBM) are established by Malik

et al. (2022), in which several hybrid ML models were generated and inspected based on the combination of supervised ML approaches. It was exposed that the hybridization of some methods produced an important benefit over the most progressive methods.

Ghosh et al. (2022) revealed a Neural Aggregate Generator (NAG) method to extract the features based on neural networks that acquire feature aggregates. The NAG's network scheme thoroughly resembles the configuration of feature aggregates, in contrast to other programmed feature extraction methods. Furthermore, the NAG improved learnable aggregates over predictable ones by easy feature value matching and relative weights. The generative adversarial networks (GAN) technique was developed by Langevin et al. (2022), in which the influence of artificial information on the performance is reliant on basic customer distributions and the source of information. In order to increase training sets for fraud detection, it has the potential to increase model performance by accumulating small quantities of GAN-generated synthetic data.

Carcillo et al. (2021) devised supervised and unsupervised approaches for the application of a hybrid strategy that uses unsupervised outlier scores to increase a fraud detection classifier's feature set. Beyond its applicability, the datasets of credit card transactions and the effect of the invention lie in the execution and assessment of several granularity levels for the purpose of an outlier score. Attention mechanism and long short-term memory (LSTM) deep model were introduced by Benchaji et al. (2021) to relate the assets of several ML approaches, like a swarm intelligence-based approach, which was utilized to select the best division of relevant features. The uniform manifold approximation and projection (UMAP) approach was used to decrease dataset dimensionality, and SMOTE was employed to address the problem of imbalanced data. To overcome the issue of imbalanced data, the prediction efficacy during the identification of fraudulent transactions was enhanced.

Lin and Jiang (2021) developed an autoencoder-probabilistic random forest (AE-PRF) to reduce the data dimensionality and to remove data features. Additionally, it used RF with a probabilistic arrangement to make data appear fraudulent and offer a probability to that label. When the related probability influences a certain threshold, the AE-PRF reports the final classification as fraudulent. Data resampling approaches, such as SMOTE, were utilized to balance the quantities of legitimate and fraudulent transactions. Strelcenia and Prakoonwit (2023) introduced a method of K CGAN, which was used to compare and assess the performance of classifiers in characteristics between criminal and permitted transactions. It is also used to test the categorization outputs and to distinguish between fraudulent and legitimate transactions. For accessibility, recall, F1 score, recall, and accuracy precision are engaged.

Seera et al. (2021) produced an ML approach to improve the learning approaches. A Malaysian financial institution utilized the definite payment card database. In order to accurately assess the effectiveness of the recognized tactic, fraud detection was identified in other parts of the world, which would be useful to collect more real-world data from other

countries. Hidden Markov models (HMMs) were established by Lucas et al. (2020) for producing features based on HMMs that permit the incorporation of sequential knowledge in transactions as features. These HMM-based characteristics allow the Random Forest classifier to categorize the information using sequential data. HMMs agree to take in a wide range of sequential data due to their numerous perspective assets.

Fiore et al. (2019) devised a GAN to tackle the class imbalance issue while using supervised classification to find credit card fraud. A training set is utilized to create an augmented set, which has more instances of the minority class than the training set. A tailored GAN is used to create synthetic instances. The advised method was fundamentally reliant on labeled examples of fraudulent transactions. Alshutayri (2023) employed logistic regression for fraud transaction detection on movie ticket transactions done using credit cards on data collected from European cardholders. The dataset consisted of 284,807 transactions, with 492 fraud transactions. The accuracy of the suggested technique was reported as 99.88%.

A parenclitic network was demonstrated by Zanin et al. (2018) to show how data mining and complex networks are combined as corresponding tools in a synergistic manner to improve the classification rates obtained by classical data mining algorithms. To advance data mining methods, fraud instances were identified in credit card transactions. Carcillo, Dal Pozzolo, et al. (2018) revealed the SCAlable Real-time Fraud Finder (SCARFF) method to identify fraud in a short period of time mechanically. The scheme, progress, and testing of an open-source big data solution for actual fraud detection were done using a comfortable real-life data set, which establishes the framework's novel contribution. In order to give emphasis to the approach, it has been entirely open source and replicable using a synthetic dataset and Docker9 container.

Carcillo, Le Borgne, et al. (2018) developed an active learning system to analyze the accuracy of CCFD. Various approaches, including semi-supervised learning, exploratory active learning, combining functions, and traditional active learning, are applied to a real-life dataset. The two-dimensional cognitive therapy of complications, such as non-separability, the issue in fraud detection, is delivered through various tactics, and it is similar to visualization. Artificial Immune Recognition System (AIRS) algorithm within artificial immune system (AIS) was devised by Halvaiee and Akbari (2014) to identify credit card fraud, and a new technique called AIS-based Fraud Detection Model (AFDM) was established for this purpose. The AIRS method has established many advancements in the technique by improving accuracy while lowering system training time and costs. An exclusive method in AFDM was utilized to regulate the antigens' empathy for one another.

Zhang et al. (2022) established a one-class support vector machine (OCSVM) and AdaBoost, in which SMOTE has resolved the imbalanced class problem with excessive performance, and their categorization capability needs to be developed in various practical circumstances. To precisely identify fraudulent activities, it utilized anomaly detection

on imbalanced data and IForest with kernel principal component analysis. OCSVM and AdaBoost approaches were active in sensing outliers, which significantly improved detection effectiveness and accuracy. Alshawi (2023) proposed an approach for detecting fraud that could deal with tiny, unbalanced datasets. He trained Logistic regression, Decision Trees, Naïve Bayes, AdaBoost, Random Forest, and XGBoost on synthetic data generated using GAN. He reported that, except for naïve Bayes, all other models had an accuracy of 95% or more.

LSTM and gated recurrent unit (GRU) were demonstrated by Mienye and Sun (2023), which serve as the source learners for a robust deep-learning policy, which also included a multilayer perceptron (MLP) functioning as the meta-learner. To balance the stability of the class distribution in the dataset, the hybrid synthetic minority oversampling model-edited nearest neighbor (SMOTE-ENN) technique is used. Du et al. (2023) introduced an autoencoder and decision tree-based classifier with Light Gradient Boosting (AED-LGB) approach, and bank credit card fraud was addressed using the AED-LGB. This method first customizes an auto-encoder to remove the features from the input, after which it feeds them into the LightGBM approach for calculation and organization. The algorithm then recognizes information that goes beyond the threshold as fake information. AED-LGB approach performance was estimated by using an anonymized database from a bank.

Hafez et al. (2025) offered a systematic review of AI-enhanced techniques used in CCFD, focusing on recent advancements in ML, DL, and hybrid models, critically evaluating over 50 studies by comparing methodologies, feature engineering techniques, and performance outcomes across different datasets. Farabi et al. (2024) also conducted a comprehensive study focused on evaluating multiple ML algorithms for CCFD, aiming to identify models that offer optimal performance in terms of accuracy and fraud mitigation. After analysis of algorithms like decision trees, random forest, SVM, logistic regression, and neural networks on benchmark datasets, the authors concluded that ensemble models, particularly random forest and gradient boosting, demonstrated superior detection capabilities.

Optimization-Based Approaches

Prabhakaran and Nedunchelian (2023) formulated the technique, oppositional cat swarm optimization-based feature selection model with a deep learning model for CCFD (OCSODL-CCFD), to identify and categorize credit card fraud. The OCSODL-CCFD method includes a variety of developments, such as oppositional cat swarm optimization (OCSO)-based feature selection, preprocessing, chaotic Krill Herd algorithm (CKHA)-based hyper-parameter optimizer, and bidirectional gated recurrent unit (BiGRU) classifier. The OCSO models' design aids in decreasing computing complexity and improving classification outcomes. Group Search Firefly Algorithm was presented by Jovanovic et al. (2022) to tackle the problem of CCFD, and it recommends a hybrid ML and swarm meta-heuristic method.

Extreme learning machines, SVMs, and extreme gradient-boosting ML methods were adjusted using the newly upgraded firefly technique called the group search firefly approach.

Padhi et al. (2022) employed the Rock Hyrax Swarm Optimization Feature Selection (RHSOFS) method to progress credit card fraud transaction documentation methods. The feature selection (FS) method in terms of a metaheuristic approach called RHSOFS, which was stimulated by the normal behavior of rock hyrax swarms. From a high-dimensional dataset, a subgroup of the best applicable features is preferred using this method. Singh and Jain (2020) presented a cost-sensitive learning flower pollination metaheuristic algorithm (CSFPA) with the benefit of the correlation-based feature selection (CFS) model and flower pollination algorithm (FPA). The other method of cost-sensitive metaheuristic technique and cost-sensitive classifier was offered to ease the misclassification cost of credit card transactions from class imbalance data.

Hyper-Heuristic Evolutionary Algorithm (HHEA), an exclusive Bayesian network classifier (BNC) approach, was established by Sánchez et al. (2009) to address a practical issue of CCFD. A HHEA method, which advances unique solutions for categorizing datasets, originally produced the FraudBNC method, and this algorithm is common enough to resolve further classification problems.

SMOTE-Based Methods

SMOTE and Easy Ensemble were introduced by Dal Pozzolo et al. (2017) to address the fraud detection issue and recommend suitable performance measures for fraud detection tasks like average precision (AP), area under curve (AUC), and precision rank. In relation to the overall number of transactions, fraud is quite occasional. Additionally, the purpose of detection is to provide the investigators with the transactions that pose the most significant risk of fraud. Therefore, rating transactions according to their chance of fraud is more important. Rtayli and Enneya (2020) developed a hyperparameter optimization (HPO) approach using SMOTE to generate a hybrid algorithm for CCFD. The advised method has a great capability to distinguish fraudulent transactions. More precisely, the Recursive Feature Elimination (RFE) was utilized for indicating the most valuable predictive features, and the Grid Search cross-validation (CV) for SMOTE and HPO are used to resolve the imbalanced data problem, and are united to produce robustness.

TL-Based Methods

Fraud detection system (FDS) was introduced by Lebichot et al. (2021) to investigate the presentation of TL methods in transaction-based fraud detection systems. It is very reasonable from a business standpoint, and the improvement of techniques to relate detection models learned in combined markets to fresh ones was still a primary goal for transactional organizations.

Other Methods

Habibpour et al. (2023) employed an uncertainty quantification (UQ) approach in which the grade of uncertainty associated with generating the predictions was measured using three deep UQ methods, which resulted in a responsible categorization. In order to advance fraud prevention, decision-makers might get supplementary insights by enumerating the uncertainty of predictable fraudulent transactions. To evaluate the prediction uncertainty estimations, various performance measures and the UQ confusion matrix are also engaged. Information utilization method (INUM) was devised by Han et al. (2021) to support four newly developed MMEAs to do a better job of handling multimodal multi-objective issues. Its inventive technique is to arbitrarily choose multiple D-vectors from all D-vectors in the choice space, to categorize them, to select the top and bottom D-vectors, to make a data vector by deducting the top from the bottom D-vectors, and finally to produce offspring by accumulating data for all solutions to produce the best results.

Prusti et al. (2021) demonstrated a graph database model in response to the graph database pattern, and a fraud detection system was offered. Some transaction database features are united with graph features that are recovered using the Neo4j tool. To precisely recognize fraudulent transactions, five supervised and two unsupervised ML approaches are applied. For detecting fraudulent transactions, the categorization methods are also utilized to calculate the features. Tingfei et al. (2020) developed a variational automatic coding (VAE) to address the issue, a VAE-based oversampling technique that syndicates traditional DL methods. In an unbalanced dataset, the VAE approach is used to generate a large number of diverse instances from minority groups, which are subsequently used to train the classification network. Additionally, it outperforms existing oversampling systems based on GAN methods.

Gianini et al. (2020) established a Game Theory-based methodology that assigns a normalized score to every individual rule, evaluating the rule's contribution to the pool's overall performance. By using the Shapley value (SV), a power index produced under Coalitional Game Theory, it was to extend the efficacy of association. This score has two major uses: to maintain or eliminate a rule from the pool during the periodic rule evaluation process and to select the best k rules. Weighted extreme learning machine (WELM) was developed by Zhu et al. (2020) to equate the performance of many intelligent optimization practices on an unbalanced classification problem while enhancing the WELM. The WELM with a probability-based mutation-enabled dandelion approach performs WELM with a bat algorithm, a dandelion algorithm, particle swarm optimization, and a genetic algorithm. The advised method is also used to detect credit card fraud and has excellent detection performance.

Darwish (2020) produced a two-level CCFD method to increase classification accuracy and accelerate detection convergence. A two-level artificial bee colony algorithm (ABC)

and k-means approaches were utilized to recognize credit card fraud from extremely unbalanced datasets. To deal with the k-means classifier's inability and to identify the precise cluster, cluster ABC was used.

Ebiaredoh-Mienye et al. (2021) presented a stacked sparse autoencoder approach integrated with an artificial neural network (ANN) to enhance the prediction accuracy of credit card fraud on high-dimensional and imbalanced data by employing feature learning techniques that automatically extract relevant patterns. Their proposed approach outperformed traditional neural networks and other ML classifiers in prediction performance while effectively reducing noise and dimensionality. Esenogho et al. (2022) integrated a neural network-based ensemble model and feature engineering techniques to improve the detection of credit card fraud, especially on imbalanced datasets. And they reported superior performance in fraud detection accuracy, F1-score, and false positive reduction compared to standalone models.

Table 1 presents a comparative analysis of various techniques reviewed in this paper on various parameters, like techniques employed, datasets used, and metrics used for evaluation.

A consumer incentive system was introduced by Wang et al. (2019) to end fraudulent credit card transactions. The two normally employed tactics are to take no preventative measures and to employ the ML detection method for all transactions. It also includes contribution motivations to customers and demanding inferior verification for every transaction, and the difficulties associated with fraudulent credit card transactions are captured by the required considerations. Somasundaram and Reddy (2019) developed a transaction window bagging (TWB) approach, which is based on incremental learning and challenges the proficient management of concept drifts caused by covariate drift

Table 1
Comparative analysis

Authors	Year	Proposed method	Dataset used	Evaluation metrics	Category
Carcillo et al.	2021	Supervised and unsupervised techniques	CCRD dataset	Precision, AUC-PR	ML
Benchaji et al.	2021	Attention mechanism and an LSTM deep model	Credit card fraud dataset	Accuracy, precision, recall	ML
Xie and Li	2021	HELMDD	Credit card transaction dataset	Recall, G-mean, AUC	Ensemble
Alharbi et al.	2022	Text2IMG conversion technique	Kaggle dataset	Accuracy, sensitivity, specificity, F1 score	DL
Alfaiz and Fati	2022	AllKNN-CatBoost	CCRD dataset	AUC, accuracy, recall, precision, F1 score	ML

Table 1 *(continue)*

Authors	Year	Proposed method	Dataset used	Evaluation metrics	Category
Malik et al.	2022	AdaBoost, LightGBM	IEEE-CIS fraud detection	ROC, TPR, precision, F-measure, TNR, Type-I error, Type-II error	ML
Jovanovic et al.	2022	Group Search Firefly Algorithm	CCRD	Accuracy, precision, recall, F1 score	Optimization
Forough and Momtazi	2022	DNN, PGM	CCRD	Precision, recall, F1 score, AUC-ROC, PR-AUC	DL
Carneiro et al.	2022	VCCA	Credit card transactions fraud detection dataset	F1 score, AUC, PRC	Clustering
Padhi et al.	2022	RHSOFS	Credit card fraud dataset	Accuracy, recall, precision, F1 score, MCC, specificity	Optimization
Zhang et al.	2022	OCSVM and AdaBoost	Credit card fraud dataset	Accuracy, precision, recall, F1 score	ML
Habibpour et al.	2023	UQ techniques	IEEE-CIS fraud detection	Accuracy, sensitivity, specificity, precision	Others
Mienye and Sun	2023	LSTM and GRU	Credit card fraud detection	Sensitivity, specificity, AUC	ML
Du et al.	2023	AED-LGB algorithm	Credit card transaction dataset	Accuracy, MCC, TPR, TN	ML
Prabhakaran and Nedunchelian	2023	OCSODL-CCFD	Credit card dataset	Accuracy, F1 score, MCC	Optimization
Karthikeyan et al.	2023	CSO-DCNN	Insurance fraud dataset	Accuracy, MAE, MSE	DL, optimization
Ebiaredoh-Mienye et al.	2021	Stacked sparse autoencoder + ANN	Credit card default dataset (likely a real-world/bank dataset)	Accuracy, precision, recall, F1 score	ML
Esenogho et al.	2022	Neural network-based ensemble model	CCRD dataset	Accuracy, F1 score, false positive rate	ML

Note. CCRD = Credit card fraud detection; ML = Machine learning; LSTM = Long short-term memory; HELMDD = Heterogeneous ensemble learning model based on data distribution; G-mean = Geometric mean; AUC = Area under curve; DL = Deep learning; AllKNN = All-k-nearest neighbors; LightGBM = Light Gradient Boosting Machine; IEEE-CIS = Institute of Electrical and Electronics Engineers – Computational Intelligence Society; ROC = Receiver operating characteristic; TPR = True positive rate; TNR = True negative rate; DNN = Deep neural network; PGM = Probabilistic graphical model; AUC-ROC = Area under the curve – Receiver operating characteristic; PR-AUC = Precision-recall area under the curve; VCCA = Value clustering for categorical attributes; PRC = Precision-recall curve; RHSOFS = Rock Hyrax Swarm Optimization Feature Selection; MCC = Matthews correlation coefficient; OCSVM = One-class support vector machine; UQ = Uncertainty quantification; GRU = Gated recurrent unit; AED-LGB = Autoencoder and decision tree-based classifier with Light Gradient Boosting; TN = True negative; OCSODL-CCFD = Oppositional cat swarm optimization deep learning for credit card fraud detection; CSO-DCNN = Competitive swarm optimization-Deep convolutional neural network; MAE = Mean absolute error; MSE = Mean squared error; ANN = Artificial neural network

and class drift. Their approach can also handle noisy and unbalanced data that creates extremely accurate predictions. To decrease the training period and offer predictions that are impervious to data imbalances, balanced data selection kept the lowest imbalance ratios.

Jiang et al. (2018) devised a fraud detection method to form a new behavioral profile of a cardholder that can be drawn on the forms of behavior from similar cardholders. To address the parameters' capacity and to respond quickly, cardholders' transactional behaviors are used. The efficiency and efficacy of the technique are confirmed by experimental findings. Advanced Transaction Exploration (APATE) was presented by Van Vlasselaer et al. (2015) for detecting fraudulent credit card transactions made at internet dealers. This method syndicated two types of features, namely intrinsic features and network-based features. The intrinsic features were a result of the characteristics of incoming transactions and customer history. Then, network-based features are derived by leveraging the network of credit card holders and merchants.

Seeja and Zareapoor (2014) developed Fraud Miner, a CCFD model designed to detect fraud in highly complex and unknown credit card transaction datasets, thereby identifying both lawful and fraudulent transaction configurations. Each client has frequent item set mining, which resolves the class imbalance problem. It was also utilized to determine if the configurations are fraudulent or legal and to make a judgment in line with the result.

RESEARCH GAPS AND ISSUES

The investigation difficulties encountered by clustering are as follows: The CBM method failed to focus on online transactions. Moreover, it was challenging to develop a new model and to create a distinct method for card-present and internet transactions. The SBS technique does not attain optimum and best outcomes, and thus, it was not easy to investigate in terms of improving the introduced approach.

The investigation challenges tackled by DL methods are revealed below: The goal of the text2IMG-based classification method is to increase the effectiveness of the method by the application of added portions. Moreover, it was stimulating to enhance prediction proficiency and to remove redundant data from transactions. The CSO-DCNN technique does not take any effort, and it was a challenge to incorporate further database, which covers statistically delicate features, and in executing ensemble attractive methods.

The experimentation problems conflicting with ML approaches are given below: The learning algorithm was unsuccessful in exchanging the linear accumulation of subsequent prospects, and it was difficult to improve the attentive precision by executing the learning-to-rank technique. AdaBoost and LightGBM techniques did not focus on the disadvantages of misplaced values. Furthermore, various approaches can be established through feature extraction and selection, allowing for the regulation of their influence on estimated accuracy.

The GAN approach was challenging to observe substitute variance in privacy explanations and allowances, which also includes perceiving the problem in privacy with the related information. Attention mechanism and LSTM deep method aim to train about CCFD approaches, in which it was difficult to process the series, which depends on transformers and attention without any recurrent networks. The AE-PRF model failed to advance its performance by utilizing the hyperparameters of the random forest (RF) and autoencoder (AE) approaches.

Moreover, there is a problem in applying the introduced technique to various requests for assessing its process. The devised model HMM did not synchronize the predictions of HMM features and LSTM, which is also a challenge to identify the scams in the face-to-face transaction. In the SCARFF technique, the challenge is to focus on analyzing the prevailing result for the business partner, assessing its effectiveness, and evaluating its robustness.

The analysis difficulties addressed by optimization are established as follows: The CSFPA did not demonstrate analysis on a larger number of datasets. Moreover, it was still a challenge to determine the time cost, and the performance of the devised technique also needs to be developed through subsequent features of unit tactics.

The difficulties encountered by the other techniques are also demonstrated below: the INUM approach aims to improve their performance by involving the challenge in contributing to real-life optimization issues. The recall rate is enhanced in the approach of VAE. It was also challenging to improve the metrics of F-score and precision, while maintaining the performance of recall equivalent to that of other techniques.

COMPARATIVE ANALYSIS

This section illuminates a discussion and analysis of CCFD based on various criteria. For this analysis, various research papers are compared on various approaches, such as characterized techniques, datasets, publication year, and performance metrics.

Analysis Based on Techniques

In this part, several CCFD techniques are compared. Figure 2 describes the distribution of various CCFD methods used in this survey. It is evident that the maximum papers surveyed used ML-based approaches, followed by optimization-related techniques. Other papers used

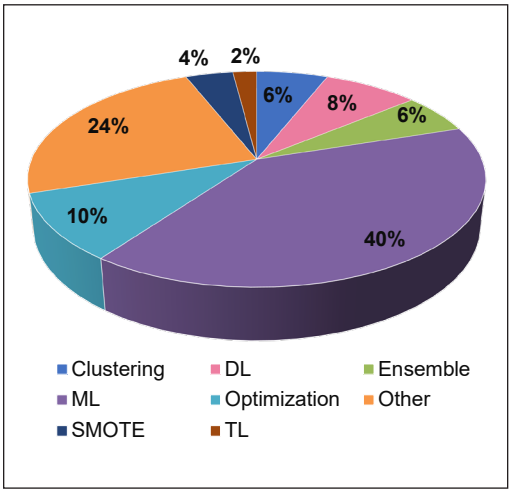


Figure 2. Category analysis based on various techniques
Note. DL = Deep learning; SMOTE = Synthetic minority oversampling technique; ML = Machine learning; TL = Transfer learning

clustering-based, DL, ensemble approaches, SMOTE techniques, and TL models for the detection of fraud.

Analysis Based on the Tools Used

In this section, the tools used by different researchers are compared. Table 2 shows the tools and software used for detecting credit card fraud. The implementation tools used by researchers are Apache, Python, Java, and MATLAB. Python and MATLAB were utilized by most of the researchers included in this survey. Apache and Java were utilized in one paper, respectively.

Table 2
Analysis based on tools

Tools	Apache	Python	Java	MATLAB
No. of publications	1	15	1	6

Analysis Based on Publication Year

Various CCFD methods were studied in this analysis. Table 3 illustrates the year of publication and the number of papers published. Most of the papers studied have been published in the years 2022 and 2023.

Table 3
Analysis based on the publication year

Year	2014	2015	2017	2018	2019	2020	2021	2022	2023
No. of publications	3	1	2	5	4	8	8	11	11

Analysis Based on the Dataset

In this section, the analysis is done on the datasets used by the papers included in the review. The datasets utilized in this analysis are BankSim dataset, Brazilian bank dataset, card-specific transaction datasets, credit card dataset, credit card fraud dataset, CCFD, credit card transaction data set, e-commerce transactions dataset, Institute of Electrical and Electronics Engineers-Computational Intelligence Society (IEEE-CIS) fraud detection, insurance fraud dataset, Kaggle dataset, PagSeguro dataset, real-world credit card transactions dataset, real-world dataset, Resilient Distributed Dataset (RDD), synthetic testing dataset, and UCSD-FICO dataset, to name a few. Figure 3 describes the datasets applied for the analysis. Here, the CCFD dataset is used in nine research papers, the credit card dataset is utilized in six papers, and a real-world dataset is used in five papers. Four papers used the credit card transactions dataset, and in three papers, the

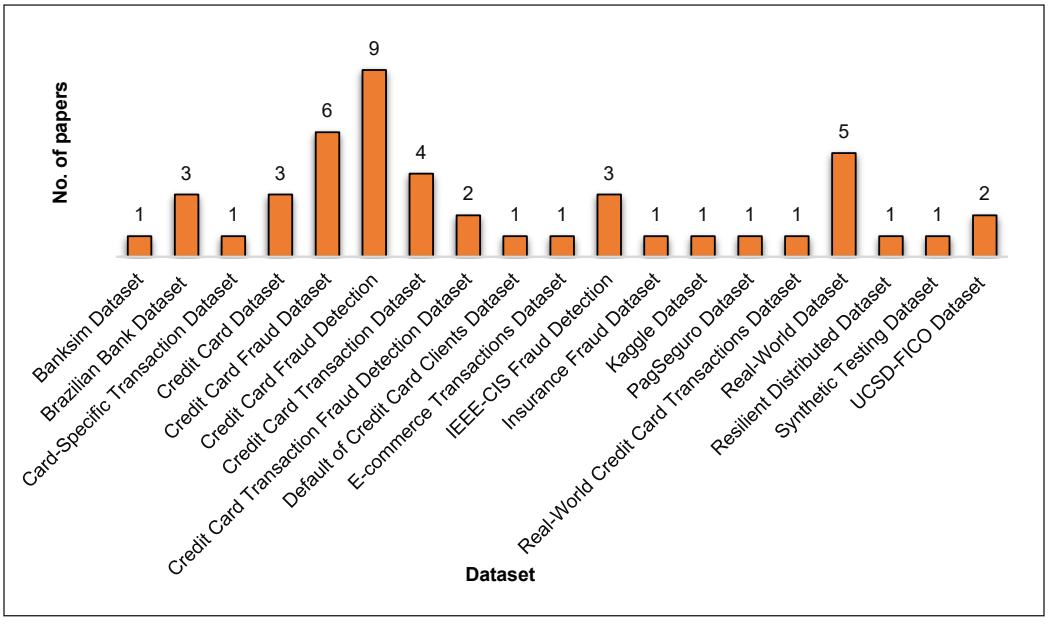


Figure 3. Analysis based on datasets
Note. IEEE-CIS = Institute of Electrical and Electronics Engineers-Computational Intelligence Society; UCSD-FICO = University of California, San Diego-Fair Isaac Corporation

Brazilian bank dataset, the credit card dataset, and the IEEE-CIS fraud detection dataset were engaged. The datasets UCSD-FICO and credit card transactions for fraud detection are used by two papers. One paper also utilizes datasets such as synthetic testing, RDD, real-world credit card transactions, PagSeguro, Kaggle, Insurance Fraud, e-commerce transactions, default of credit card clients, BankSim, and card-specific transaction datasets. The datasets used have many attributes; some are used in their original form, while others are used after feature extraction. Analysis done on these datasets can give us insight into what features generally distinguish fraud transactions from normal transactions. ML and DL algorithms were able to perform efficiently on real-world datasets as well as datasets available online. The research done on different datasets can be used by banking and credit card companies to identify fraudulent transactions and prevent fraud, leading to a decrease in losses due to fraud.

Analysis Based on Performance Metrics

In this analysis section, the evaluation metrics applied in the research papers are explained. The performance metrics utilized in this dataset are accuracy, precision, recall, F measure, specificity, AUC, sensitivity, Matthew's correlation coefficient (MCC), true positive rate (TPR), true negative rate (TNR), false positive rate (FPR), cost, Type II error, Type I error, receiver operating characteristic- area under curve (ROC-AUC score), receiver

operating characteristic (ROC), root mean quadratic error (RMQE), precision-recall curve (PRC), precision at k (Pk), negative predictive value (NPV), mean square error (MSE), misclassification rate, absolute mean error (AME), Kolmogorov–Smirnov (K-S statistics), hit rate, G-mean, Fraud detection rate, false negative rate (FNR), false alarm rate, fall-out, error rate, detection rate, process capability index (CPk), benefit-cost ratio (BCR), average cost, area under ROC curve (AUROC), area under the precision-recall (AUPR measure), area under the receiver operating characteristic (AUC-ROC), area under the curve of precision-recall (AUCPR), alert rate, and AP. Here, accuracy was used in 26 papers, precision in 25 papers, F-measure in 19 papers, recall in 21 papers, AUC in 11 papers, specificity in 12 papers, TPR and FPR in six papers, TNR in five papers, AUC-ROC, AUC-PR, BCR, and G-mean in two papers and finally, AP, alert rate, AUPR measure, AUROC, average cost, CPk, detection rate, error rate, fall-out, false alarm rate, FNR, fraud detection rate, hit rate, K-S statistics, MAE, misclassification rate, MSE, NPV, Pk, PRC, RMQE, ROC, ROC-AUC score, Type I and II errors are identified in only one paper, respectively.

Figure 4 illuminates the metrics-based analysis. As seen from the graph, accuracy is most frequently used, followed by precision, recall, and F-measure. Accuracy provides an overall evaluation of the method employed, including both fraudulent and normal transactions. Precision, on the other hand, emphasizes measuring the proportion of true positive predictions among all positive predictions. Recall measures the proportion of true positive predictions among all actual positive instances in the dataset. Recall is important in detecting fraud. High recall indicates fewer fraud transactions remain unidentified.

Analysis Based on Values of Accuracy

The range of accuracy and the number of research papers are given in Table 4, which indicates that accuracy ranges between the values of 0-80, 81-85, 86-99, and above 99%, respectively. Here, the accuracy between 0-80% lies in two papers. Next, between the values of 81-85%, two papers were selected. From 16 papers, the accuracy has a value between 86-99% and six research papers reached an accuracy range of above 99%.

Table 4
Analysis based on accuracy

Accuracy range (%)	Number of research papers
0-80	2
81-85	2
86-99	16
Above 99	6

CONCLUSION

This review aims to study past work in the field of CCFD. The survey focused on techniques employing clustering-, ML-, optimization-, SMOTE-, TL-, ensemble, and DL-based methods, respectively. Furthermore, the challenges in the literature are considered and identified from past research. Class imbalance is found in most of the CCFD datasets. Some

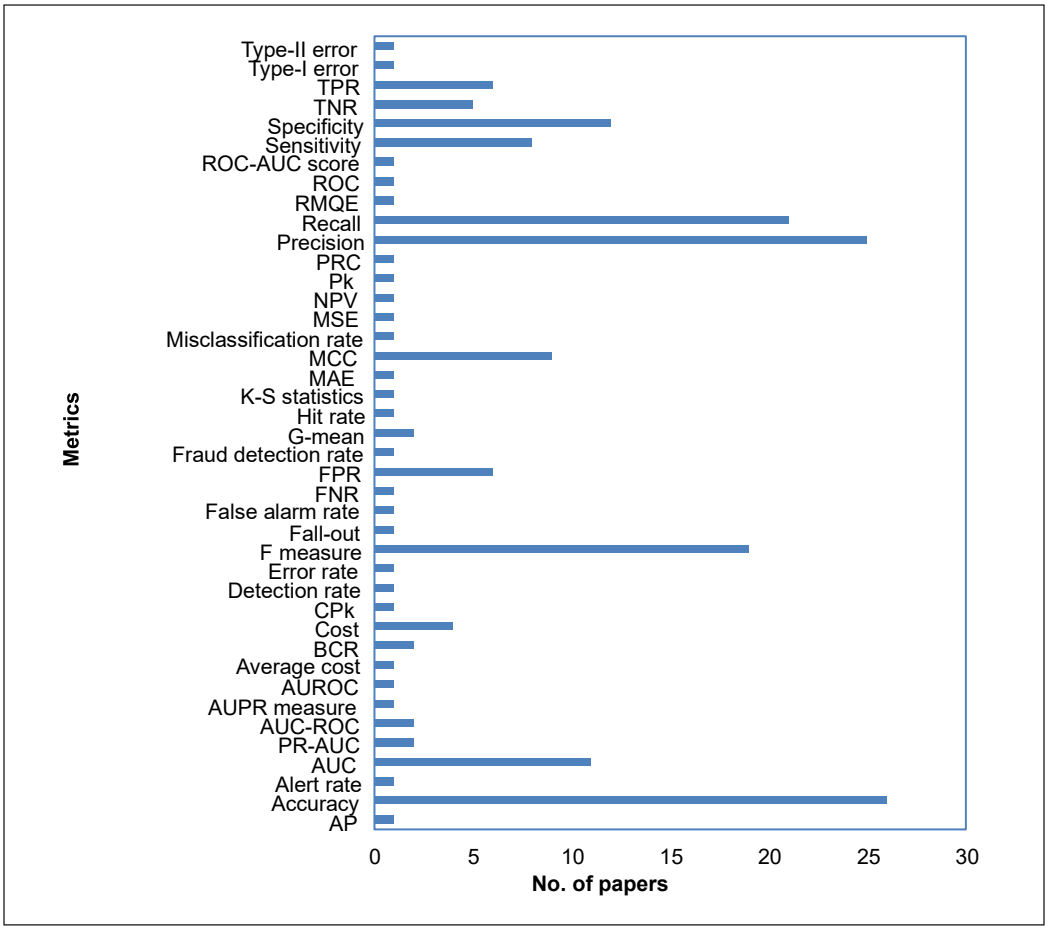


Figure 4. Analysis based on performance metrics

Note. TPR = True positive rate; TNR = True negative rate; ROC-AUC = Receiver operating characteristic-area under the curve; ROC = Receiver operating characteristic; RMQE = Root mean quadratic error; PRC = Precision-recall curve; Pk = Precision at k; NPV = Negative predictive value; MSE = Mean squared error; MCC = Matthews correlation coefficient; MAE = Mean absolute error; K-S statistics = Kolmogorov-Smirnov statistics; G-mean = Geometric mean ; FPR = False positive rate; FNR = False negative Rate; CPk = Process capability index; BCR = Balanced classification rate; AUROC = Area under the receiver operating characteristic curve; AUPR = Area under the precision-recall; AUC-ROC = Area under the curve–Receiver operating characteristic; PR-AUC = Precision-recall area under the curve; AUC = Area under curve; AP = Average precision

techniques that have been used for reducing the class imbalance problem are also discussed. Moreover, we have tried to identify the research gaps in the techniques analyzed. Different techniques are analyzed on various parameters. The analysis is conducted on the tools used for analysis, the datasets employed, the year of publication of papers, performance metrics, and other relevant factors. In this survey, it was discovered that Python was used more frequently by the researchers for analysis, followed by MATLAB. To assess performance,

accuracy was employed. The database used for this analysis was taken from online sources and some real-world datasets. In the future, research can be conducted to address the class imbalance problem present in CCFD datasets, and feature extraction techniques can be employed to reduce the processing time. The research should aim to enhance accuracy and speed in CCFD, enabling banks and other organizations to benefit from these techniques.

ACKNOWLEDGEMENTS

The authors would like to thank Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India for their academic support throughout the completion of this research.

REFERENCES

Afriyie, J. K., Tawiah, K., Pels, W. A., Addai-Henne, S., Dwamena, H. A., Owiredu, E. O., Ayeh, S. A., & Eshun, J. (2023). A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. *Decision Analytics Journal*, *6*, 100163. https://doi.org/10.1016/j.dajour.2023.100163

Ahmad, H., Kasasbeh, B., Aldabaybah, B., & Rawashdeh, E. (2023). Class balancing framework for credit card fraud detection based on clustering and similarity-based selection (SBS). *International Journal of Information Technology*, *15*, 325–333. https://doi.org/10.1007/s41870-022-00987-w

Alfaiz, N. S., & Fati, S. M. (2022). Enhanced credit card fraud detection model using machine learning. *Electronics*, *11*(4), 662. https://doi.org/10.3390/electronics11040662

Alharbi, A., Alshammari, M., Okon, O. D., Alabrah, A., Rauf, H. T., Alyami, H., & Meraj, T. (2022). A novel text2IMG mechanism of credit card fraud detection: A deep learning approach. *Electronics*, *11*(5), 756. https://doi.org/10.3390/electronics11050756

Alhashmi, A. A., Alashjaee, A. M., Darem, A. A., Alanazi, A. F., & Effghi, R. (2023). An ensemble-based fraud detection model for financial transaction cyber threat classification and countermeasures. *Engineering, Technology and Applied Science Research*, *13*(6), 12433–12439.

Alshawi, B. (2023). Utilizing GANs for credit card fraud detection: A comparison of supervised learning algorithms. *Engineering, Technology and Applied Science Research*, *13*(6), 12264–12270.

Alshutayri, A. (2023). Fraud prediction in movie theater credit card transactions using machine learning. *Engineering, Technology and Applied Science Research*, *13*(3), 10941-10945.

Bahnsen, A. C., Aouada, D., & Ottersten, B. (2015). Example-dependent cost-sensitive decision trees. *Expert Systems with Applications*, *42*(19), 6609–6619. https://doi.org/10.1016/j.eswa.2015.04.042

Benchaji, I., Douzi, S., El Ouahidi, B., & Jaafari, J. (2021). Enhanced credit card fraud detection based on attention mechanism and LSTM deep model. *Journal of Big Data*, *8*, 151. https://doi.org/10.1186/s40537-021-00541-8

Carcillo, F., Dal Pozzolo, A., Le Borgne, Y.-A., Caelen, O., Mazzer, Y., & Bontempi, G. (2018). *SCARFF*: A scalable framework for streaming credit card fraud detection with Spark. *Information Fusion*, *41*, 182–194. https://doi.org/10.1016/j.inffus.2017.09.005

Carcillo, F., Le Borgne, Y.-A., Caelen, O., & Bontempi, G. (2018). Streaming active learning strategies for real-life credit card fraud detection: Assessment and visualization. *International Journal of Data Science and Analytics*, *5*, 285–300. https://doi.org/10.1007/s41060-018-0116-z

Carcillo, F., Le Borgne, Y.-A., Caelen, O., Kessaci, Y., Oblé, F., & Bontempi, G. (2021). Combining unsupervised and supervised learning in credit card fraud detection. *Information Sciences*, *557*, 317–331. https://doi.org/10.1016/j.ins.2019.05.042

Carneiro, E. M., Forster, C. H. Q., Mialaret, L. F. S., Dias, L. A. V., & da Cunha, A. M. (2022). High-cardinality categorical attributes and credit card fraud detection. *Mathematics*, *10*(20), 3808. https://doi.org/10.3390/math10203808

Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2017). Credit card fraud detection: A realistic modeling and a novel learning strategy. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(8), 3784–3797. https://doi.org/10.1109/TNNLS.2017.2736643

Dal Pozzolo, A., Caelen, O., Le Borgne, Y.-A., Waterschoot, S., & Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. *Expert Systems with Applications*, *41*(10), 4915–4928. https://doi.org/10.1016/j.eswa.2014.02.026

Dal Pozzolo, A., Johnson, R., Caelen, O., Waterschoot, S., Chawla, N. V., & Bontempi, G. (2014). Using HDDT to avoid instances propagation in unbalanced and evolving data streams. In *International Joint Conference on Neural Networks* (pp. 588–594). IEEE. https://doi.org/10.1109/IJCNN.2014.6889638

Darwish, S. M. (2020). An intelligent credit card fraud detection approach based on semantic fusion of two classifiers. *Soft Computing*, *24*, 1243–1253. https://doi.org/10.1007/s00500-019-03958-9

Du, H., Lv, L., Guo, A., & Wang, H. (2023). AutoEncoder and LightGBM for credit card fraud detection problems. *Symmetry*, *15*(4), 870. https://doi.org/10.3390/sym15040870

Ebiaredoh-Mienye, S. A., Esenogho, E., & Swart, T. G. (2020). Integrating enhanced sparse autoencoder-based artificial neural network technique and Softmax regression for medical diagnosis. *Electronics*, *9*(11), 1963. https://doi.org/10.3390/electronics9111963

Ebiaredoh-Mienye, S. A., Esenogho, E., & Swart, T. G. (2021). Artificial neural network technique for improving prediction of credit card default: A stacked sparse autoencoder approach. International *Journal of Electrical and Computer Engineering*, *11*(5), 4392–4402. http://doi.org/10.11591/ijece.v11i5.pp4392-4402

Ebiaredoh-Mienye, S. A., Swart, T. G., Esenogho, E., & Mienye, I. D. (2022). A machine learning method with filter-based feature selection for improved prediction of chronic kidney disease. *Bioengineering*, *9*(8), 350. https://doi.org/10.3390/bioengineering9080350

Esenogho, E., Mienye, I. D., Swart, T. G., Aruleba, K., & Obaido, G. (2022). A neural network ensemble with feature engineering for improved credit card fraud detection. *IEEE Access*, *10*, 16400–16407. https://doi.org/10.1109/ACCESS.2022.3148298

Farabi, S. F., Prabha, M., Alam, M., Hossan, M. Z., Arif, M., Islam, M. R., Uddin, A., Bhuiyan, M., & Biswas, M. Z. A. (2024). Enhancing credit card fraud detection: A comprehensive study of machine learning algorithms and performance evaluation. *Journal of Business and Management Studies*, *6*(3), 252–259. https://doi.org/10.32996/jbms.2024.6.13.21

Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Information Sciences*, *479*, 448–455. https://doi.org/10.1016/j.ins.2017.12.030

Forough, J., & Momtazi, S. (2022). Sequential credit card fraud detection: A joint deep neural network and probabilistic graphical model approach. *Expert Systems*, *39*(1), e12795. https://doi.org/10.1111/exsy.12795

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, *46*(4), 1–37. https://doi.org/10.1145/2523813

Ghosh, K. D., Jurgovsky, J., Siblini, W., & Granitzer, M. (2022). NAG: Neural feature aggregation framework for credit card fraud detection. *Knowledge and Information Systems*, *64*, 831–858. https://doi.org/10.1007/s10115-022-01653-0

Gianini, G., Fossi, L. G., Mio, C., Caelen, O., Brunie, L., & Damiani, E. (2020). Managing a pool of rules for credit card fraud detection by a game theory based approach. *Future Generation Computer Systems*, *102*, 549–561. https://doi.org/10.1016/j.future.2019.08.028

Habibpour, M., Gharoun, H., Mehdipour, M., Tajally, A., Asgharnezhad, H., Shamsi, A., Khosravi, A., & Nahavandi, S. (2023). Uncertainty-aware credit card fraud detection using deep learning. *Engineering Applications of Artificial Intelligence*, *123*(Part A), 106248. https://doi.org/10.1016/j.engappai.2023.106248

Hafez, I. Y., Hafez, A. Y., Saleh, A., Abd El-Mageed, A. A., & Abohany, A. A. (2025). A systematic review of AI-enhanced techniques in credit card fraud detection. *Journal of Big Data*, *12*, 6. https://doi.org/10.1186/s40537-024-01048-8

Halvaiee, N. S., & Akbari, M. K. (2014). A novel model for credit card fraud detection using artificial immune systems. *Applied Soft Computing*, *24*, 40–49. https://doi.org/10.1016/j.asoc.2014.06.042

Han, S., Zhu, K., Zhou, M., & Cai, X. (2021). Information-utilization-method-assisted multimodal multiobjective optimization and application to credit card fraud detection. *IEEE Transactions on Computational Social Systems*, *8*(4), 856–869. https://doi.org/10.1109/TCSS.2021.3061439

Jain, Y., Tiwari, N., Dubey, S., & Jain, S. (2019). A comparative analysis of various credit card fraud detection techniques. *International Journal of Recent Technology and Engineering*, *7*(5S2), 402–407.

Jiang, C., Song, J., Liu, G., Zheng, L., & Luan, W. (2018). Credit card fraud detection: A novel approach using aggregation strategy and feedback mechanism. *IEEE Internet of Things Journal*, *5*(5), 3637–3647. https://doi.org/10.1109/JIOT.2018.2816007

Jovanovic, D., Antonijevic, M., Stankovic, M., Zivkovic, M., Tanaskovic, M., & Bacanin, N. (2022). Tuning machine learning models using a group search firefly algorithm for credit card fraud detection. *Mathematics*, *10*(13), 2272. https://doi.org/10.3390/math10132272

Karthik, V. S. S., Mishra, A., & Reddy, U. S. (2022). Credit card fraud detection by modeling behavior pattern using hybrid ensemble model. *Arabian Journal for Science and Engineering*, *47*, 1987–1997. https://doi.org/10.1007/s13369-021-06147-9

Karthikeyan, T., Govindarajan, M., & Vijayakumar, V. (2023). An effective fraud detection using competitive swarm optimization based deep neural network. *Measurement: Sensors*, *27*, 100793. https://doi.org/10.1016/j.measen.2023.100793

Kim, E., Lee, J., Shin, H., Yang, H., Cho, S., Nam, S. K., Song, Y.-K., Yoon, J.-A., & Kim, J.-I. (2019). Champion-challenger analysis for credit card fraud detection: Hybrid ensemble and deep learning. *Expert Systems with Applications*, *128*, 214–224. https://doi.org/10.1016/j.eswa.2019.03.042

Kültür, Y., & Çağlayan, M. U. (2017). A novel cardholder behavior model for detecting credit card fraud. In *9th International Conference on Application of Information and Communication Technologies* (pp. 148–152). IEEE. https://doi.org/10.1109/ICAICT.2015.7338535

Langevin, A., Cody, T., Adams, S., & Beling, P. (2022). Generative adversarial networks for data augmentation and transfer in credit card fraud detection. *Journal of the Operational Research Society*, *73*(1), 153–180. https://doi.org/10.1080/01605682.2021.1880296

Lebichot, B., Verhelst, T., Le Borgne, Y.-A., He-Guelton, L., Oblé, F., & Bontempi, G. (2021). Transfer learning strategies for credit card fraud detection. *IEEE Access*, *9*, 114754–114766. https://doi.org/10.1109/ACCESS.2021.3104472

Li, Z., Liu, G., & Jiang, C. (2020). Deep representation learning with full center loss for credit card fraud detection. *IEEE Transactions on Computational Social Systems*, *7*(2), 569–579. https://doi.org/10.1109/TCSS.2020.2970805

Lin, T.-H., & Jiang, J.-R. (2021). Credit card fraud detection with auto-encoder and probabilistic random forest. *Mathematics*, *9*(21), 2683. https://doi.org/10.3390/math9212683

Lucas, Y., Portier, P.-E., Laporte, L., He-Guelton, L., Caelen, O., Granitzer, M., & Calabretto, S. (2020). Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs. *Future Generation Computer Systems*, *102*, 393–402. https://doi.org/10.1016/j.future.2019.08.029

Mahmoudi, N., & Duman, E. (2015). Detecting credit card fraud by modified Fisher Discriminant Analysis. *Expert Systems with Applications*, *42*(5), 2510–2516. https://doi.org/10.1016/j.eswa.2014.10.037

Malik, E. F., Khaw, K. W., Belaton, B., Wong, W. P., & Chew, X. (2022). Credit card fraud detection using a new hybrid machine learning architecture. *Mathematics*, *10*(9), 1480. https://doi.org/10.3390/math10091480

Mienye, I. D., & Sun, Y. (2023). A deep learning ensemble with data resampling for credit card fraud detection. *IEEE Access*, *11*, 30628–30638. https://doi.org/10.1109/ACCESS.2023.3262020

Obaido, G., Ogbuokiri, B., Swart, T. G., Ayawei, N., Kasongo, S. M., Aruleba, K., Mienye. I. D., Arubela, I., Chukwu, W., Osaye, F., Egbelowo, O. F., Simphiwe, S., & Esenogho, E. (2022). An interpretable machine learning approach for hepatitis B diagnosis. *Applied Sciences*, *12*(21), 11127. https://doi.org/10.3390/app122111127

Padhi, B. K., Chakravarty, S., Naik, B., Pattanayak, R. M., & Das, H. (2022). RHSOFS: Feature selection using the rock hyrax swarm optimization algorithm for credit card fraud detection system. *Sensors*, *22*(23), 9321. https://doi.org/10.3390/s22239321

Prabhakaran, N., & Nedunchelian, R. (2023). Oppositional cat swarm optimization-based feature selection approach for credit card fraud detection. *Computational Intelligence and Neuroscience*, *2023*, 2693022. https://doi.org/10.1155/2023/2693022

Prusti, D., Das, D., & Rath, S. K. (2021). Credit card fraud detection technique by applying graph database model. *Arabian Journal for Science and Engineering*, *46*, 1–20. https://doi.org/10.1007/s13369-021-05682-9

Rtayli, N., & Enneya, N. (2020). Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization. *Journal of Information Security and Applications*, *55*, 102596. https://doi.org/10.1016/j.jisa.2020.102596

Sánchez, D., Vila, M. A., Cerda, L., & Serrano, J. M. (2009). Association rules applied to credit card fraud detection. *Expert Systems with Applications*, *36*(2), 3630–3640. https://doi.org/10.1016/j.eswa.2008.02.001

Seeja, K. R., & Zareapoor, M. (2014). FraudMiner: A novel credit card fraud detection model based on frequent itemset mining. *The Scientific World Journal*, *2014*, 252797. https://doi.org/10.1155/2014/252797

Seera, M., Lim, C. P., Kumar, A., Dhamotharan, L., & Tan, K. H. (2021). An intelligent payment card fraud detection system. *Annals of Operations Research*, *334*, 445–467. https://doi.org/10.1007/s10479-021-04149-2

Singh, A., & Jain, A. (2020). Cost-sensitive metaheuristic technique for credit card fraud detection. *Journal of Information and Optimization Sciences*, *41*(6), 1319–1331. https://doi.org/10.1080/02522667.2020.1809090

Somasundaram, A., & Reddy, S. (2019). Parallel and incremental credit card fraud detection model to handle concept drift and data imbalance. *Neural Computing and Applications*, *31*, 3–14. https://doi.org/10.1007/s00521-018-3633-8

Strelcenia, E., & Prakoonwit, S. (2023). Improving classification performance in credit card fraud detection by using new data augmentation. *AI*, *4*(1), 172–198. https://doi.org/10.3390/ai4010008

Tingfei, H., Guangquan, C., & Kuihua, H. (2020). Using variational auto encoding in credit card fraud detection. *IEEE Access*, *8*, 149841–149853. https://doi.org/10.1109/ACCESS.2020.3015600

Van Vlasselaer, V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Snoeck, M., & Baesens, B. (2015). *APATE*: A novel approach for automated credit card transaction fraud detection using network-based extensions. *Decision Support Systems*, *75*, 38-48. https://doi.org/10.1016/j.dss.2015.04.013

Wang, D., Chen, B., & Chen, J. (2019). Credit card fraud detection strategies with consumer incentives. *Omega*, *88*, 179–195. https://doi.org/10.1016/j.omega.2018.07.001

Xie, Y., Li, A., Gao, L., & Liu, Z. (2021). A heterogeneous ensemble learning model based on data distribution for credit card fraud detection. *Wireless Communications and Mobile Computing*, *2021*, 2531210. https://doi.org/10.1155/2021/2531210

Zanin, M., Romance, M., Moral, S., & Criado, R. (2018). Credit card fraud detection through parenclitic network analysis. *Complexity*, *2018*, 5764370. https://doi.org/10.1155/2018/5764370

Zhang, Y.-F., Lu, H.-L., Lin, H.-F., Qiao, X.-C., & Zheng, H. (2022). The optimized anomaly detection models based on an approach of dealing with imbalanced dataset for credit card fraud detection. *Mobile Information Systems*, *2022*, 8027903. https://doi.org/10.1155/2022/8027903

Zhu, H., Liu, G., Zhou, M., Xie, Y., Abusorrah, A., & Kang, Q. (2020). Optimizing weighted extreme learning machines for imbalanced classification and application to credit card fraud detection. *Neurocomputing*, *407*, 50–62. https://doi.org/10.1016/j.neucom.2020.04.078

Pertanika J. Sci. & Technol. 33 (6): 2665 - 2683 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Impact of Different Drone Sprayer Nozzle Types on Bio-Fertilizer Application in Wetland Paddy Cultivation

Emmanuel Otache Adanu[1,8], Ahmad Suhaizi Mat Su[2,3,4]*, Azmi Yahya[1,3,4], Siti Amni Ismail[5], Nik Norasma Che'Ya[2], Muhammad Huzaifah Mohd Roslim[6], Abdul Shukor Juraimi[7], Muhamad Noor Hazwan Abd Manaf[7] and Mst. Motmainna[7,9]

[1]Department of Biological and Agricultural Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

[2]Department of Agriculture Technology, Faculty of Agriculture, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

[3]Smart Farming Technology Research Centre, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

[4]Institute Plantation Studies, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

[5]Faculty of Plantation and Agrotechnology, Universiti Teknologi MARA Melaka, Jasin Campus, 77300 Merlimau, Melaka, Malaysia

[6]Department of Crop Science, Faculty of Agricultural Science and Forestry, Universiti Putra Malaysia Bintulu Campus, 97000 Bintulu, Sarawak, Malaysia

[7]Department of Crop Science, Faculty of Agriculture, Universiti Putra Malaysia, 43400 Serdang, Malaysia

[8]Department of Agriculture Education, Federal College of Education (Technical), P.M.B. 60, Gombe, Nigeria

[9]Laboratory of Climate-Smart Food Production, Institute of Tropical Agriculture, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

ABSTRACT

The effectiveness and efficiency of chemical spray applications on cropland are strongly influenced by the size of the nozzle. This research study focuses on examining the spray characteristics associated with two different nozzle sizes using bio-fertilizer. The experimentation took place on a 13.85-ha paddy field. To collect spray deposits, water-sensitive papers (WSPs) were strategically placed in a setup consisting of 39 WSPs arranged at distances of 0.5 m apart in three rows spaced 1 m apart. An agricultural spraying drone was deployed at a speed of 3 m/s and a height of 2 m to conduct the tests. Parametric data, essential for both descriptive and inferential statistical analyses, were generated using the DepositScan software. The results of the characteristic test indicated that approximately 50% of the collected droplet sizes were below 89.11±20.59 and 239±78.44

ARTICLE INFO

Article history:
Received: 02 January 2025
Accepted: 17 April 2025
Published: 08 October 2025

DOI: https://doi.org/10.47836/pjst.33.6.12

E-mail addresses:
engr.adanu@gmail.com (Emmanuel Otache Adanu)
asuhaizi@upm.edu.my (Ahmad Suhaizi Mat Su)
azmiy@upm.edu.my (Azmi Yahya)
sitiamni@gmail.com (Siti Amni Ismail)
niknorasma@upm.edu.my (Nik Norasma Che'Ya)
muhammadhuzaifah@upm.edu.my (Muhammad Huzaifah Mohd Roslim)
ashukur@upm.edu.my (Abdul Shukor Juraimi)
mnor901@gmail.com (Muhamad Noor Hazwan Abd Manaf)
motmainna@upm.edu.my (Mst. Motmainna)
* Corresponding author

© Universiti Putra Malaysia Press

μm for the fine and coarse nozzle sizes, respectively. The coefficients of variation were calculated as 0.27 for the very fine nozzle and 0.85 for the coarse nozzle, indicating varying levels of droplet size uniformity. The *t*-test analysis revealed a significant difference ($P \leq 0.05$) between the two nozzle sizes for all investigated parameters, except for the drift collected within a 30-m distance from the target. Specifically, the fine nozzle type produced very fine droplets with superior spray uniformity, while the coarse nozzle types generated moderate to large droplet sizes. This highlights the critical role of selecting the appropriate nozzle in determining the overall quality of spray applications.

Keywords: Climate action, droplet size, foliar fertilizer, Malaysia, spray distribution, spray drift, zero hunger

INTRODUCTION

The adoption of drone technology in agriculture has revolutionized farming practices by enhancing precision, efficiency, and sustainability. Unmanned aerial vehicles (UAVs), commonly known as drones, are increasingly used for crop monitoring, field mapping, and aerial spraying, offering a promising alternative to conventional, labor-intensive methods (Ahirwar et al., 2019; Nhamo et al., 2020; Tsouros et al., 2019). Equipped with advanced cameras, sensors, and spraying systems, drones enable farmers to apply fertilizers and pesticides with unprecedented accuracy, reducing waste and environmental impact. The characteristics of the spray, influenced by factors such as nozzle type, pressure, and environmental conditions, play a critical role in determining droplet size, drift potential, and deposition uniformity, all of which directly impact crop health and yield (Hunter et al., 2019; Ismail et al., 2021; Taylor et al., 2004).

In Malaysia, drone technology for fertilization and pesticide spraying is gaining traction, mainly through service providers in key rice granaries like Tanjung Karang and Sekinchan, as well as large plantations. Studies have highlighted its advantages (Dorairaj & Govender, 2023; Rosedi & Shamsi, 2022), such as the reduction of labor by 75%, cost savings of RM49.15/ha, and increased efficiency (2.4 ha/hr vs. 0.7 ha/hr manually). Adoption is driven by labour shortages, 52.5% faster operations, and sustainability goals, backed by government initiatives such as the Civil Aviation Authority of Malaysia (CAAM)'s free-fly zones and training programs, including My Drone Services. However, barriers persist, such as high costs, dependence on service providers (70%), regulatory hurdles, and skill gaps, with 53.7% of farmers having only average drone knowledge (Harun et al., 2024). While drones enhance precision, reduce chemical exposure, and attract youth to agriculture, self-operated use remains challenging in resource-limited rural areas.

In response to the environmental challenges associated with chemical fertilizers, such as soil degradation, water pollution, and greenhouse gas emissions, bio-fertilizers have emerged as a sustainable alternative. Unlike conventional fertilizers, bio-fertilizers use beneficial microorganisms to enhance nutrient availability, support soil biodiversity,

and promote healthier crop growth (Bahadur et al., 2006; Iqbal et al., 2020). This eco-friendly approach reduces dependency on chemical inputs, aligning with global efforts like the European Union's target to reduce chemical fertilizer use by 50% by 2030 (Montanarella & Panagos, 2021). In addition to improving soil health, bio-fertilizers can also contribute to climate resilience and biodiversity conservation, making them a promising solution for sustainable agriculture (Bumandalai & Tserennadmid, 2019; Youssef & Eissa, 2014).

Rice, a staple food for over half of the world's population, particularly in Asia, South America, and parts of Africa, faces increasing demand and production challenges (Ghani et al., 2024; Papademetriou, 2000). In Malaysia, for example, rice production fell from over 1.85 million metric tons in the year 2014 to 1.44 million metric tons in 2023, as self-sufficiency remains elusive due to fluctuating per-hectare yields and reliance on imported rice (Economic Planning Unit [EPU], 2015; Siddharta, 2025). As global demand for agricultural products is projected to increase by 60% by 2030, sustainable practices are essential to ensure food security and environmental health (Montanarella & Panagos, 2021). The integration of UAV technology for bio-fertilizer application in rice cultivation holds significant potential to address these challenges by enhancing productivity and reducing environmental impacts.

Selecting the correct nozzle size is crucial for optimizing fertilizer application efficiency and maximizing crop yield. It ensures precise application rates, preventing over- or under-application, which can lead to nutrient runoff, environmental pollution, or nutrient deficiencies. A properly sized nozzle provides uniform coverage, minimizing waste and ensuring consistent crop growth, as demonstrated in studies on spray pattern uniformity (Matthews et al., 2014; Zhu et al., 2011). It also optimizes droplet size for better nutrient absorption, whether through leaf uptake in foliar applications (de Castro & Schjoerring, 2024) or soil penetration for liquid fertilizers (Niemoeller et al., 2011). By reducing fertilizer drift and evaporation (Chethan et al., 2019), the right nozzle selection enhances efficiency, lowers input costs, and improves overall yield. This precision is particularly critical for modern practices, such as drone-based applications, where nozzle choice directly impacts droplet deposition accuracy and environmental sustainability (Han et al., 2025; P. Wang et al., 2024).

This study examines the effects of bio-fertilizer spray characteristics on paddy cultivation using drone sprayers equipped with different nozzle sizes. By comparing the spray deposition, drift potential, and droplet size achieved with various nozzle configurations, this research aims to identify optimal spray parameters for effective bio-fertilizer application in rice fields. This information is crucial for guiding farmers and policymakers on the appropriate use of UAV technology in precision agriculture, ultimately supporting more sustainable and efficient rice production systems.

MATERIALS AND METHODS

Test Site

The test was conducted on 20th August 2020 in a 13.85-ha paddy field at Kg. Padang Raja Melor, Kota Bharu, Kelantan, Malaysia (05°57'43''N, 102°17'21''E), managed by the Kemubu Agricultural and Development Authority (KADA) (Figure 1). The age of paddy was 54 days after transplanting (DAT) with an average height of 68 cm, and a water level of 20-30 cm was maintained.

The test took place between 9:00 and 11:00 a.m. to ensure minimal and stable wind conditions and avoid direct overhead sun. Weather data, including temperature, relative humidity, wind speed and direction, and altitude, were collected before the experiment using a handheld anemometer (SKU-H9619, Kkmoon, China). Measurements were taken approximately 2 m above the ground, in the upwind direction, for less than 5 s. The experimental site experienced an average temperature of 32.1±1.9°C and a relative humidity of 65.10±12.8%. Wind speed was measured at 2.05±0.8 m/s, predominantly blowing from the North-West direction. The experiment was conducted at an altitude of 5.4 m above sea level. The recommended relative humidity should be above 50% and wind speed below 5 m/s to prevent treatment volatilization, as suggested by C. Wang et al. (2023) and Lodwik et al. (2020).

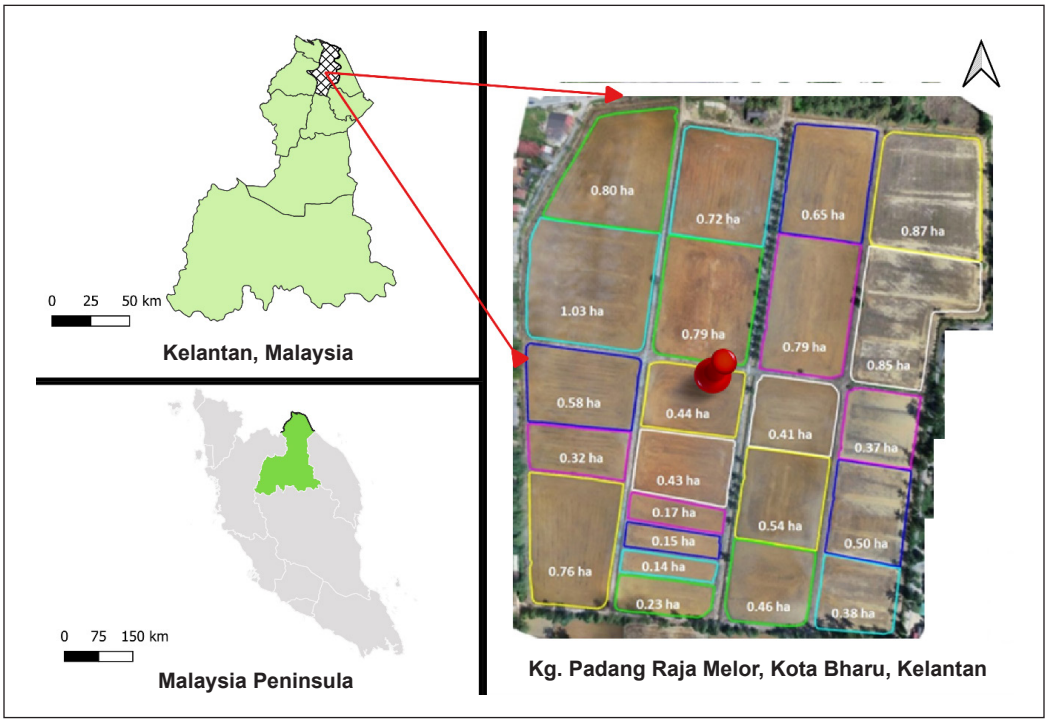


Figure 1. Map of the paddy field

Spray Systems

The spray system for the test was a Pola V-16s agriculture spraying drone (Poladrone Solutions Sdn. Bhd., Malaysia) (Figure 2). The drone is equipped with a spray system situated underneath it for precision spray control and features an extended spray boom with a trapezoidal geometry, where the rear nozzles are more closely spaced than the fore pair. The fore and rear nozzle pairs are controlled separately by two compatible pumps, providing the system with three spraying modes. The spray system on the unmanned aircraft system (UAS) responds to commands and spray operations accurately and effectively. The drone is equipped with a Global Positioning System (GPS) real-time kinematics (RTK) differential system, which offers precise differential positioning and is capable of recording flight parameters on the go. The technical parameters of the UAS are presented in Table 1.

Table 1
Unmanned aircraft system technical parameters

Classification	Parameters
Motors	6
Total weight (kg)	35.5
Dimension (Extended) (mm)	2,364.6 × 2,144.6 × 556
Battery capacity (mAh)	22,000 × 2 units
Maximum runtime (min)	20
Operation method	Remote control/ Autonomous
Height above crop canopy (m)	1.5 - 10
Number of nozzles	4
Unit flow (L/min)	0.6 – 2.0
Load capacity (L)	16
Spraying width (m)	1.5 - 3.5

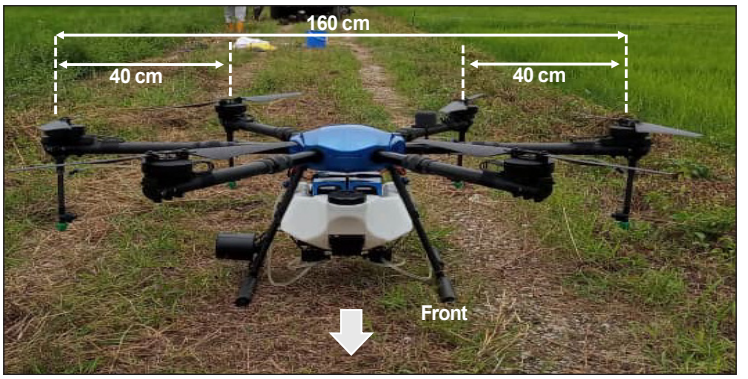


Figure 2. Drone configuration with four nozzles beneath each rear rotor, spaced 40 cm apart

Spraying Test

The nozzles used for the spray tests were the Teejet® XR11008VS and XR110015VS (Spraying Systems Co., USA) flat fan nozzle types (Table 2). They spray at a maximum angle of 110° and have operational parts made of steel for increased durability. The XR11008VS produces a medium-coarse spray (color code: white), while the XR110015VS nozzle produces a fine spray (color code: green).

Actual flow rate tests were conducted for both nozzles by collecting the water from the nozzles in a conical flask at a system application rate of 20 L/ha for 1 min. The spray liquid was a 10 ml/L solution of bio-fertilizer in water. For each nozzle size, the tests were replicated three times to validate the results. Each nozzle was calibrated prior to the spraying operation, with average unit flow rates of 0.927±0.003 and 0.530±0.00 L/min for the white and green nozzles, respectively (Table 2). The spray nozzle tips were made of stainless steel, chosen to be used during the test for being more durable and less prone to wear out due to excessive liquid pressure and heavy utilization.

The spray treatment used for the tests was Booster Grow 8+ bio-fertilizer (DNG Worldwide, Malaysia), enriched with nitrogen, phosphorus, potassium, and beneficial microbes to enhance plant growth, improve soil fertility, and strengthen disease resistance. The recommended application rate ranges from 2.5–10 ml/L, depending on the crop type, including tree crops, field crops, and vegetables. In this study, a 10 ml/L solution in water was prepared at a total volume of 20 L and applied using a UAS at a constant rate of 20 L/ha, with a mean travel speed of 3 m/s. WSPs (Chongqing Liuliushanxia Plant Protection Technology Co., Ltd., China) measuring 30 mm × 25 mm were precut from their original 80 mm × 30 mm size for droplet collection and analysis.

The WSPs were arranged on a frame standing at the height of the rice plants and spanning 6m in length. Three frames, spaced 5 m apart and arranged parallel to the direction of the prevailing wind, constituted a unit test. The WSPs were positioned on the frame with a total coverage swath width of 6 m at 0.5 m intervals, with the zero (0) mark at the center of the WSP arrangement and the negative parts upwind. A setup was made for drift collection in both directions of the swath at 1.5, 6.5, 10.5, and 28.5 m from the swath path,

Table 2
Drone flying and spraying parameters and nozzle specification

Spray volume (L/ha)	20	
Spray height (m)	2.00±0.010	
Horizontal UAS speed (m/s)	3.00±0.10	
Total number of nozzles	4	
Spraying angle	110°	
Nozzle type	XR11008VS	XR110015VS
Nozzle color	White	Green
Calibrated unit flowrate (L/min)	0.927±0.003	0.530±0.000
Droplet size classification	Medium	Fine
Nozzle type with stainless steel tips		

Note. UAS = Unmanned aircraft system

with the outer rows spread at 30° from the mid row to accommodate the unpredictability of the wind. This entire setup, as shown in Figure 3, was replicated three times under each flight mission, spaced at 60 m apart. The UAS flowed across the WSP arrangement, passing through the zero mark while spraying the bio-fertilizer mix at a constant height and speed (Figures 4 and 5). The details of the UAS spray operation are found in Table 2. The spray deposits were allowed a few minutes to settle before the sample papers were collected using disposable gloves. The samples were carefully collected into labelled plastic office file jackets and later scanned using an HP ScanJet Pro 2500 f1 Flatbed Scanner (HP Inc., USA). The images were scanned at an optical resolution of 600 dpi and saved in JPEG file format.

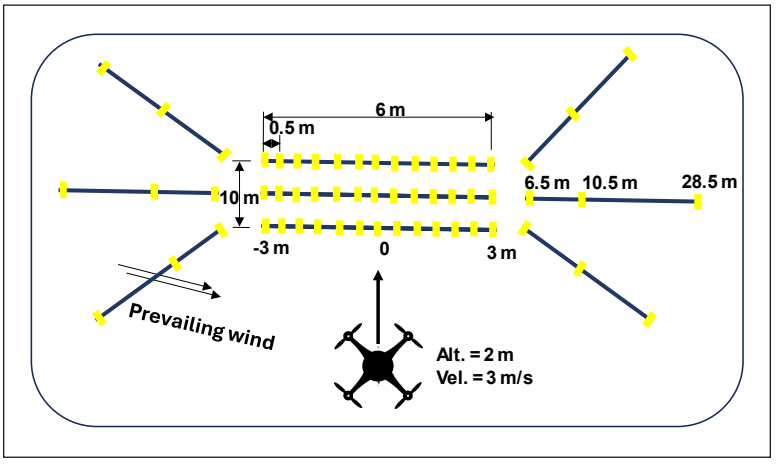


Figure 3. Experimental field layout
Note. Alt. = Altitude; Vel. = Velocity



Figure 4. Unmanned aircraft system spraying test operation in the paddy cultivation

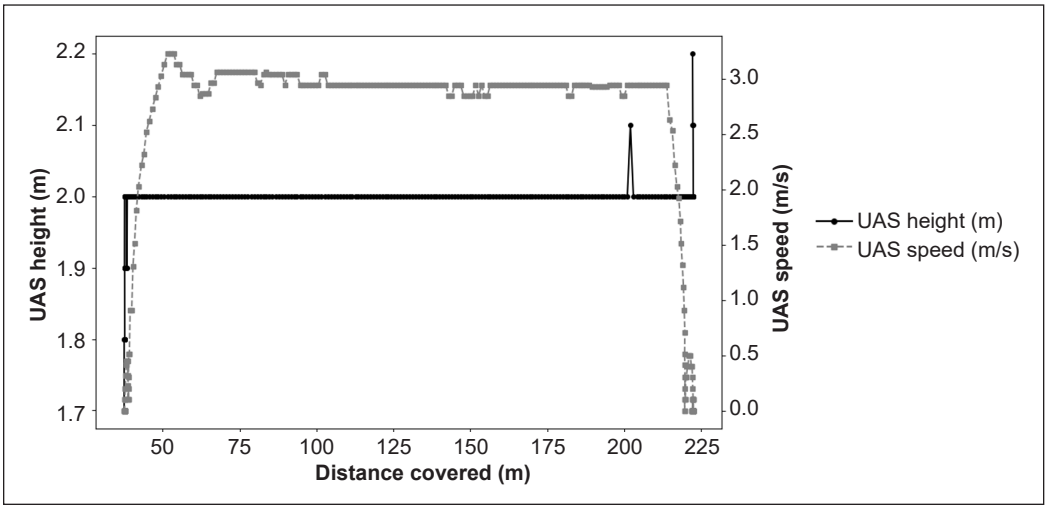


Figure 5. Recorded unmanned aircraft system (UAS) flight data showing stages (take-off, hovering, and spraying) and height status during the spraying operation

Determination of Effective Spray Swath

To determine the effective spray swath, a deposition threshold limit was selected for the UAS flight and spray settings during the test. A deposition volume per area of 0.2 µl/cm² was chosen as the limit. The end-to-end sampling points with a deposition volume of at least 0.2 µl/cm² on both sides of the route were defined as the starting and ending points of the effective swath area. The distance between these two points was considered the effective swath width.

Data and Statistical Analysis

The scanned documents were analyzed using DepositScan™ software (United States Department of Agriculture [USDA], USA) (Zhu et al., 2011), which calculates Dv0.1, Dv0.5, Dv0.9, percentage area, deposition, and droplet density. These metrics, derived from droplet size distributions, are essential for evaluating spray nozzle performance. Dv0.1, the 10^{th} percentile volume diameter, represents smaller droplets (finer droplets, more prone to drift but better coverage), Dv0.5 indicates the median droplet size (larger values mean coarser sprays), and Dv0.9, the 90^{th} percentile volume diameter, reflects larger droplets (less drift but potentially uneven coverage). In this study, spray performances were compared based on droplet density, deposition, and droplet size distributions, providing insights into nozzle efficiency and optimizing them for applications like minimizing waste or maximizing coverage.

Descriptive and inferential statistics were computed on the parametric data collected for all nozzle types after testing for normality. A *t*-test was used to test for significance

between the green and white nozzles' parametric data collected at a 95% confidence level. The coefficient of variation (CV) was estimated for both nozzle types (American Society of Agricultural and Biological Engineers [ASABE], 2012). The CV is valuable in estimating the droplet distribution uniformity within sampled points. A smaller CV value indicates better uniformity of deposited droplets. It is estimated using Equations 1 and 2:

$$CV = \frac{S}{\bar{x}} \times 100\% \quad [1]$$

While,

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \quad [2]$$

where S is the standard deviation. x_i denotes deposition at sample points, and $\bar{x}$ is the mean deposition of the total sampled points, while n refers to the number of sampled points.

RESULTS AND DISCUSSION

Table 3 summarizes the mean and standard deviation of several spray characteristics for two nozzle types, XR11008VS (white) and XR110015VS (green), measured at different points across the spray width. Key parameters include droplet size distribution (Dv0.1, Dv0.5, and Dv0.9), spray volume (µl/cm²), coverage percentage, and droplet deposition (Dep/cm²). The white nozzle produces larger droplets, with mean values for Dv0.1, Dv0.5, and Dv0.9 at 149, 261, and 363 µm, respectively, while the green nozzle averages 62.6, 90.2, and 128.4 µm, producing relatively smaller droplet sizes. The white nozzle delivers a higher mean spray volume per square centimeter, about 88.2% more than the green nozzle (white nozzle = 3.01 µl/cm² and green nozzle = 0.354 µl/cm²).

The green nozzle achieves a higher mean coverage percentage (16.7%) compared to the white nozzle (7.25%), reflecting a finer mist that spreads more uniformly across surfaces. The droplet deposition rate also shows a clear difference: the green nozzle has a significantly higher deposition rate (1,248 droplets per cm²) than the white nozzle (204 droplets per cm²), indicating a denser distribution of droplets.

Droplet Size Distribution

The finer nozzle (green code) with a relatively lower CV value shows consistent and uniform spray compared to the coarse nozzle (white code). Therefore, a green nozzle is suggested to be more precise in delivering the bio-fertilizer applications in paddy cultivation, minimizing waste and improving effectiveness. The comparison of droplet size distribution for XR11008VS (white) and XR110015VS (green) nozzles reveals significant

differences in spray quality, impacting their applications (Table 3). The white nozzle produces larger droplets, with mean values for Dv0.1, Dv0.5, and Dv0.9 at 149, 261, and 363 μm, respectively, while the green nozzle averages 62.6, 90.2, and 128.4 μm (Table 4). T-test results confirm these differences are statistically significant, indicating that the white nozzle generates substantially larger droplets than the green nozzle (Table 5). Additionally, the CV for the white nozzle is 85%, reflecting greater droplet size variability, which can enhance surface coverage. However, this high variability may also indicate inconsistencies due to nozzle wear, pressure fluctuations, or improper calibration, potentially affecting deposition uniformity. Conversely, the green nozzle's CV of 27% represents more consistent droplet sizes, resulting in a finer, more even spray pattern classified as "very fine". Ensuring proper nozzle calibration and pressure regulation is essential to minimize variability and optimize spray efficiency.

These findings imply that each nozzle is suited to different spraying requirements. The larger and more variable droplet sizes produced by the white nozzle make it well-suited

Table 3
Mean and standard deviation (SD) of spray deposition for all tested parameters of XR11008VS and XR110015VS nozzles

White nozzle type (XR11008VS)						
Point (m)	**Dv0.1 (μm)**	**Dv0.5 (μm)**	**Dv0.9 (μm)**	**Vol. D (μl/cm²)**	**Coverage (%)**	**Dep. D (Dep/cm²)**
-1.5	139±8.99	187±8.74	298±9.33	1.25±0.13	5.97±0.28	112±5.19
-1.0	152±9.97	276±12.90	299±9.34	1.23±0.02	5.79±0.26	147±6.76
-0.5	161±10.40	267±12.50	400±12.50	4.39±0.44	6.34±0.30	167±7.73
0	198±12.80	343±16.10	449±14	5.45±0.56	12±0.54	298±13.8
0.5	155±9.98	286±13.40	379±11.80	5.28±0.54	6.04±0.28	162±7.54
1.0	107±6.91	312±14.60	417±13	1.25±0.03	8.89±0.45	201±9.26
1.5	133±8.55	158±7.37	296±9.25	2.21±0.23	5.74±0.27	177±8.15
Mean±SD	149±49	261±103	363±86	3.01±2.44	7.25±4.75	204±4
Green nozzle type (XR110015VS)						
Point (m)	**Dv0.1 (μm)**	**Dv0.5 (μm)**	**Dv0.9 (μm)**	**Vol. D (μl/cm²)**	**Coverage (%)**	**Dep. D (Dep/cm²)**
-1.5	66±6.09	68±15.50	70±24.40	0.20±0.06	14.20±2.92	1,115±111
-1.0	51.30±4.74	66±15.10	83±29	0.26±0.07	14.40±2.95	1,126±112
-0.5	62.20±5.74	96.10±21.90	137±47.70	0.30±0.08	15.10±3.10	1,283±128
0	64±5.91	81.30±18.60	111±38.70	0.43±0.12	16.10±3.30	1,355±135
0.5	63.90±5.89	109±24.80	190±66.40	0.42±0.12	20.90±4.28	1,388±139
1.0	70±6.46	114±26	177±61.60	0.49±0.14	22.30±4.58	1,350±135
1.5	61±5.63	97.10±22.20	131±45.80	0.38±0.11	14±2.88	1,116±111
Mean±SD	62.60±11.30	90.20±24.20	128.40±61.60	0.35±0.15	16.70±5.60	1,248±140

Note. Dep. D = Deposition density; Vol. D = Volumetric deposition

Table 4
The t*-test result for all paired parametric data for both nozzles*

	Parameter					
	Dv0.1 (µm)	**Dv0.5 (µm)**	**Dv0.9 (µm)**	**Vol. D (µl/cm²)**	**Coverage (%)**	**Dep. D (Dep/cm²)**
t-value	7.63	6.95	11.06	2.77	-6.97	-27.62
p-value	0.00026***	0.00044***	0.00003***	0.032**	0.00044***	0.00001***

Note. Dep. D = Deposition density; Vol. D = Volumetric deposition**;** ** = 0.05 significant level; *** = 0.01 significant level

Table 5
Result for droplet size distribution

Nozzle type	Dv0.1 (µm)	Dv0.5 (µm)	Dv0.9 (µm)	CV (%)	Spray quality*
White	149±49	261±103	363±86	85	Medium
Green	62.62±11.30	90.20±24.20	128.40±61.60	27	Very fine

Note. CV = Coefficient of variation; ***** Droplet classification standards for > 3 bars (ASABE, 2009)

for applications needing minimized drift and focused spraying of bio-fertilizer in targeted areas. Meanwhile, the green nozzle's fine and uniform droplet pattern is ideal for achieving consistent coverage over broader areas, beneficial in foliar applications. However, the green nozzle's smaller droplets could be prone to drifting in windy conditions, requiring careful management to ensure optimal efficacy. The findings for XR11008VS and XR110015VS align with previous literature that identifies larger droplets as beneficial for reducing drift and improving deposition on target areas, particularly for pesticide applications (Miller & Ellis, 2000). However, these larger droplets may reduce coverage uniformity compared to finer sprays, as observed in our study. On the other hand, finer sprays, like those generated by XR110015VS, are associated with higher deposition rates and better coverage (Nuyttens et al., 2007). Nonetheless, fine sprays have a higher drift potential, particularly under windy conditions (Ferguson et al., 2016). Figure 6, displaying droplet size distribution across a 6 m swath width, demonstrates how finer droplets can achieve high coverage but are susceptible to environmental factors such as wind. Therefore, selecting a nozzle type should balance spray quality, coverage needs, and environmental considerations to optimize application effectiveness (Pascuzzi et al., 2017; Salyani et al., 2013).

Droplet Deposition Density

Table 3 and Figure 7 present data on deposition density (Dep/cm²) for the white (XR11008VS) and green (XR110015VS) nozzles across various sampling points. The green nozzle exhibits a consistently higher deposition density, peaking around 1,400 Dep/cm², particularly between the -0.5 to 1.0 m sampling points. This indicates that the green

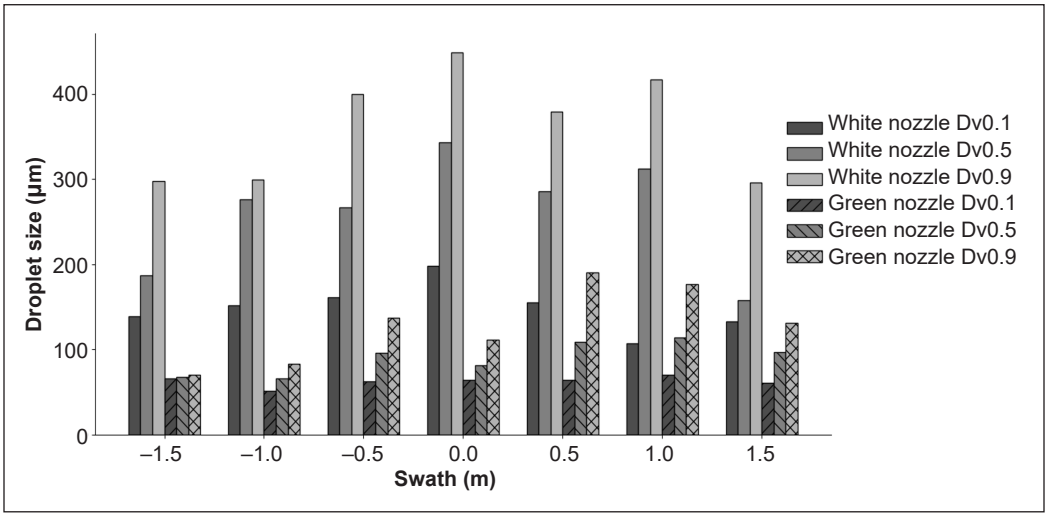


Figure 6. Droplet size distribution across a 6 m swath at 0.5 m intervals for all nozzles at Dv0.1, Dv0.5, and Dv0.9, respectively

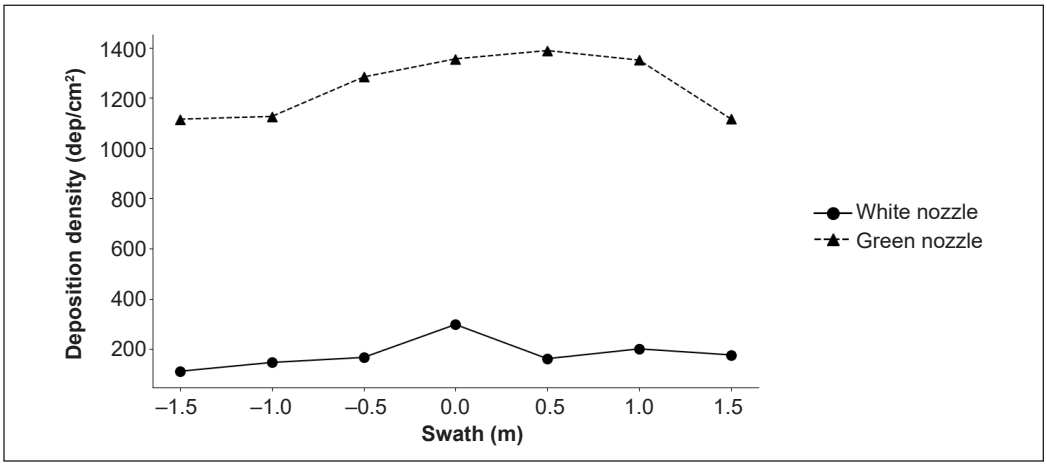


Figure 7. Mean deposit density

nozzle achieves greater deposition density across the swath width, which may enhance coverage and pesticide efficacy. In contrast, the white nozzle has a lower deposition density, fluctuating between approximately 100 and 300 Dep/cm², with its peak near the center point (0 m) at around 300 Dep/cm². Table 3 provides further detail, showing the green nozzle has a much higher average deposition density of 1,248±140 Dep/cm² compared to the white nozzle's 204±94 Dep/cm². This significant difference, supported by the *t*-test results in Table 4 (t = -27.62, $p < 0.01$), highlights the green nozzle's superior performance in deposition density. However, this comes with the trade-off of potentially increased drift due to finer droplets, as suggested by other studies (Grella et al., 2023; Nuyttens et al., 2007).

The results align with existing research on nozzle performance and droplet deposition density, crucial for effective pesticide application. Higher deposition density, as observed with the green nozzle (XR110015VS), can enhance coverage and improve pesticide efficacy by ensuring a more uniform application on target surfaces. Recent studies indicate that nozzles producing finer droplets typically yield higher deposition densities, improving coverage but potentially increasing drift risk due to smaller droplet size (Shan et al., 2021; Srinivasarao et al., 2021). However, finer droplets, while improving target coverage, can lead to off-target movement, particularly in windy conditions (Grella et al., 2023). In contrast, the white nozzle (XR11008VS) produces coarser droplets with lower deposition density, which can improve precision by reducing drift (Shan et al., 2021). This suggests that the white nozzle may be more suitable for applications where drift reduction is prioritized over maximum coverage density. The *t*-test in Table 4 confirms the significant difference between the nozzles' deposition densities, supporting the green nozzle's use in situations where maximum coverage is essential (Srinivasarao et al., 2021).

Percentage Coverage

Table 3 and Figure 8 reveal that the green nozzle (XR110015VS) consistently achieves a higher percentage coverage than the white nozzle (XR11008VS) across all sampling points. Specifically, the green nozzle reaches peak coverage at around 22% at the 1-m mark, while the white nozzle's coverage tops out at approximately 12% at 0 m. This suggests that the finer droplets from the green nozzle deliver more extensive and uniform coverage compared to the coarser droplets produced by the white nozzle. This finding is consistent with Zhang et al. (2024), who noted that finer droplets typically enhance target coverage due to their larger surface area, which improves adhesion on leaf surfaces. Such coverage is particularly advantageous when uniform pesticide distribution is essential for effectiveness.

Hunter et al. (2019) also found that equipment speed influences coverage, with the best results at lower speeds. At a speed of 7 m/s, they recorded coverage of 22, 16, and 13% for Air Induction Extended Range (AIXR), Extended Range (XR) flat-fan, and the Turbo TeeJet® Induction (TTI) nozzles, respectively. Similarly, flow rate impacts coverage; Nansen et al. (2015) observed up to 10% coverage at flow rates between 30 and 40 L/ha, indicating that both speed and flow rate adjustments can enhance coverage in field applications.

However, the increased coverage with finer droplets has a downside. Nuyttens et al. (2007) emphasized the heightened drift potential of fine droplets, especially in windy conditions. The green nozzle's higher coverage percentage, while beneficial for thorough application, could increase the risk of off-target drift compared to the white nozzle, which produces coarser droplets that offer better drift control. Thus, the recommendation on the timing for the spraying application is crucial to avoid the drift or off-target spray droplets.

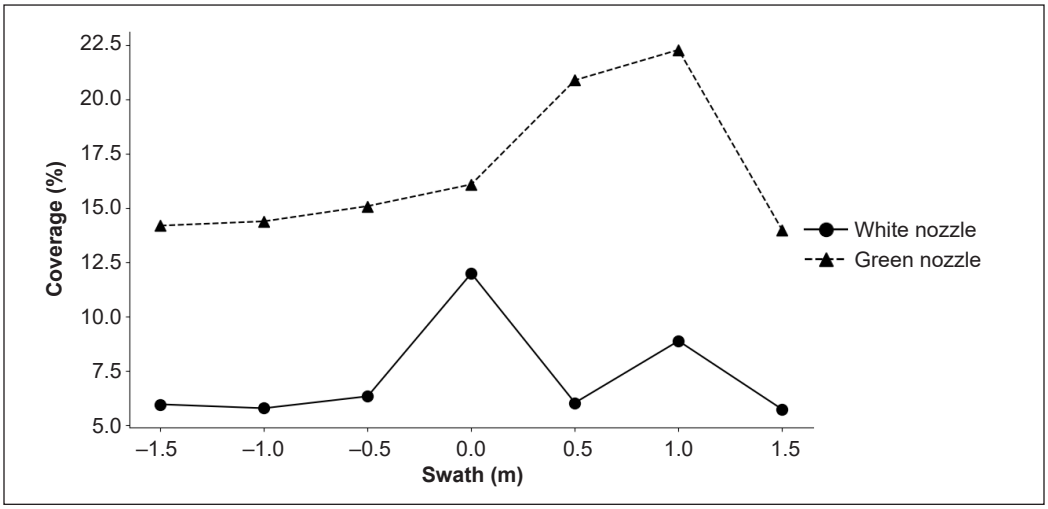


Figure 8. Mean percentage coverage from both nozzles

Drift

The ISO 22866:2005 standard (International Organization for Standardization [ISO], 2005) recommends measuring spray drift up to 50 m downwind from the swath edge. However, studies such as Xue et al. (2014) found that 90% of the drift was captured within 8 m when a UAS operated at 5 m height under 2 m/s wind conditions. This study and the results presented in Table 6 highlight the differences in spray drift deposition between the green and white nozzles at varying distances from the swath edge, along with the corresponding meteorological conditions during each test run. At the closest measurement point (1.5 m), the white nozzle deposited significantly more spray (1.60 µl/cm²) than the green nozzle (0.49 µl/cm²) (t = -9.093, p = 0.001). This suggests that the white nozzle's coarser droplets settle more rapidly, reducing the likelihood of drift under moderate wind conditions (2.05 m/s). The significantly higher deposition at this near-field location indicates its suitability for applications requiring localized spray coverage. However, as the distance increased, the green nozzle exhibited greater deposition than the white nozzle. At 6.5 m, the green nozzle deposited 0.258 µl/cm², whereas the White nozzle deposited 0.165 µl/cm² (t = 5.305, p = 0.006). By 10.5 m, this trend became more pronounced, with the green nozzle (0.140 µl/cm²) significantly outperforming the white nozzle (0.026 µl/cm²) (t = 11.779, p = 0.000). The higher drift potential of the green nozzle is likely due to its finer droplets, which remain airborne for longer durations, increasing the likelihood of mid-range deposition. At 20.5 m, both nozzles exhibited reduced deposition, but the green nozzle still showed slightly higher values (0.0161 µl/cm²) than the white nozzle (0.0090 µl/cm²) (t = 4.041, p = 0.016). This indicates that even at long distances, the finer droplets from the green nozzle continue to contribute to drift, reinforcing its potential for broader area coverage.

Table 6
Average volumetric deposition of spray drift and standard deviation (in brackets) from spray nozzles along distance, and the metrological information

Distance (m)	Volumetric deposition (µl/cm²) and standard deviation		*t*-value	*p*-value
	Green nozzle	White nozzle		
1.5	0.490 (0.023)	1.600 (0.210)	-9.093	0.001
6.5	0.258 (0.011)	0.165 (0.030)	5.305	0.006
10.5	0.140 (0.016)	0.026 (0.008)	11.779	0.000
20.5	0.016 (0.003)	0.009 (0.001)	4.041	0.016
Wind speed (m/s)	1.400 (0.500)	2.700 (0.300)	-3.862	0.018
Temperature (°C)	30.700 (1.500)	33.500 (1.000)	-2.690	0.055
R/H (%)	65.500 (7.400)	60.000 (10.000)	0.671	0.537

Note. R/H = Relative humidity

These findings are consistent with Baetens et al. (2007), who found that smaller droplets tend to drift farther, and with Grella et al. (2017), who reported that nozzle type plays a critical role in drift variability.

Meteorological conditions also varied between the test runs. Wind speed increased from 1.4 m/s during the green nozzle test to 2.7 m/s during the white nozzle test (t = -3.862, p = 0.018). Additionally, temperature increased from 30.7 to 33.5°C (t = -2.690, p = 0.055), while relative humidity dropped from 65.5 to 60.0% (t = 0.671, p = 0.537). While these changes may have had some influence on spray drift, the overall trends suggest that nozzle design remained the dominant factor affecting deposition patterns. The finer droplets of the green nozzle were more susceptible to increased wind speed, which explains their greater deposition at mid-to-long distances.

For UAS-based bio-fertilizer application, these findings emphasize the need to balance targeted deposition and drift control. The white nozzle is more effective for near-field spraying, where rapid droplet settlement is desirable, while the green nozzle is better suited for mid-range deposition, enhancing overall spray coverage. Optimizing application efficiency may require the use of adjuvants to improve deposition, real-time meteorological monitoring to adjust spraying parameters dynamically (Jomantas et al., 2023), and careful nozzle selection to minimize drift while ensuring uniform crop coverage.

CONCLUSION

The study compares two nozzles, XR11008VS (white) and XR110015VS (green), for spraying bio-fertilizer from a UAS, focusing on droplet size, deposition density, coverage, and spray drift. The white nozzle produces larger droplets (mean Dv0.5 of 261 µm) with a "Medium" spray quality, suitable for drift reduction and targeted delivery. In contrast, the green nozzle generates finer droplets (mean Dv0.5 of 90.2 µm) with a "very fine"

spray quality, ideal for achieving uniform coverage. The green nozzle achieves a higher deposition density (mean of 1,248 Dep/cm²) but increases drift risk, especially under windy conditions, while the white nozzle's coarser droplets reduce drift potential but result in lower deposition density. Drift analysis shows that the white nozzle achieves higher near-field deposition, whereas the green nozzle's finer droplets exhibit a more gradual drift pattern, leading to similar deposition at extended distances.

Overall, although the white nozzle is effective for minimizing drift and targeted spraying, the green nozzle is preferable for paddy cultivation, where uniform and thorough bio-fertilizer coverage is critical for rice plant growth and yield. With proper management of environmental conditions to control drift, the green nozzle's ability to cover larger areas can optimize input usage, lower costs, and enhance overall agronomic and economic outcomes for farmers. Additionally, practical considerations such as lower maintenance requirements, cost-effectiveness, and compatibility with existing UAS systems further reinforce its suitability for modern rice farming operations.

ACKNOWLEDGEMENTS

The authors are grateful to the Long-Term Research Grant Scheme (LRGS/1/2019/UPM01/2/5; vote number: 5545002) and the Fundamental Research Grant Scheme (FRGS/1/2018/WAB01/UPM/02/2; vote number: 5540087), and sincerely acknowledge the Universiti Putra Malaysia for providing facilities. The authors received funding from the Research Project entitled «Pest and Disease Monitoring Using Artificial Intelligence for Risk Management of Rice Under Climate Change» under the LRGS, Ministry of Higher Education, Malaysia.

REFERENCES

Ahirwar, S., Swarnkar, R., Bhukya, S., & Namwade, G. (2019). Application of drone in agriculture. *International Journal of Current Microbiology and Applied Sciences*, *8*(1), 2500–2505. https://doi.org/10.20546/ijcmas.2019.801.264

American Society of Agricultural and Biological Engineers. (2009). *ASAE/ASABE S572.1: Spray nozzle classification by droplet spectra*. ASABE.

American Society of Agricultural and Biological Engineers. (2012). *ASAE/ASABE S327.4: Terminology and definitions for application of crop or forestry production and protection agents*. ASABE.

Baetens, K., Nuyttens, D., Verboven, P., De Schampheleire, M., Nicolai, B., & Ramon, H. (2007). Predicting drift from field spraying by means of a 3D computational fluid dynamics model. *Computers and Electronics in Agriculture*, *56*(2), 161–173. https://doi.org/10.1016/j.compag.2007.01.009

Bahadur, A., Singh, J., Singh, K. P., Upadhyay, A. K., & Rai, M. (2006). Effect of organic amendments and bio-fertilizers on growth, yield and quality attributes of Chinese cabbage (*Brassica pekinensis*). *The Indian Journal of Agricultural Sciences*, *76*(10), 596–598.

Bumandalai, O., & Tserennadmid, R. (2019). Effect of *Chlorella vulgaris* as a biofertilizer on germination of tomato and cucumber seeds. *International Journal of Aquatic Biology*, *7*(2), 95–99. https://doi.org/10.22034/ijab.v7i2.582

Chethan, C. R., Singh, P. K., Dubey, R. P., Chander, S., & Ghosh, D. (2019). Herbicide application methodologies: Influence of nozzle selection, droplet size and spray drift on effective spraying – A review. *Innovative Farming*, *4*(1), 45–53.

de Castro, S. A. Q., & Schjoerring, J. K. (2024). Foliar nitrogen and phosphorus fertilization. In D. L. Sparks (Ed.), *Advances in agronomy* (Vol. 187, pp. 109–169). Academic Press. https://doi.org/10.1016/bs.agron.2024.05.002

Dorairaj, D., & Govender, N. T. (2023). Rice and paddy industry in Malaysia: Governance and policies, research trends, technology adoption and resilience. *Frontiers in Sustainable Food Systems*, *7*, 1093605. https://doi.org/10.3389/fsufs.2023.1093605

Economic Planning Unit. (2015). *Eleventh Malaysia Plan, 2016-2020: Anchoring growth on people*. https://ekonomi.gov.my/sites/default/files/2021-05/Chapter%201.pdf

Ferguson, J. C., Chechetto, R. G., Hewitt, A. J., Chauhan, B. S., Adkins, S. W., Kruger, G. R., & O'Donnell, C. C. (2016). Assessing the deposition and canopy penetration of nozzles with different spray qualities in an oat (*Avena sativa* L.) canopy. *Crop Protection*, *81*, 14–19. https://doi.org/10.1016/j.cropro.2015.11.013

Ghani, M. A. A., Juraimi, A. S., Su, A. S. M., Ahmad-Hamdani, M. S., Mominul Islam, A. K. M., & Motmainna, M. (2024). Chemical weed control in direct-seeded rice using drone and mist flow spray technology. *Crop Protection*, *184*, 106853. https://doi.org/10.1016/j.cropro.2024.106853

Grella, M., Gallart, M., Marucco, P., Balsari, P., & Gil, E. (2017). Ground deposition and airborne spray drift assessment in vineyard and orchard: The influence of environmental variables and sprayer settings. *Sustainability*, *9*(5), 728. https://doi.org/10.3390/su9050728

Grella, M., Maffia, J., Dinuccio, E., Balsari, P., Miranda-Fuentes, A., Marucco, P., & Gioelli, F. (2023). Assessment of fine droplets (<10 μm) in primary airborne spray drift: A new methodological approach. *Journal of Aerosol Science*, *169*, 106138. https://doi.org/10.1016/j.jaerosci.2023.106138

Han, L., Wang, Z., He, M., Liu, Y., & He, X. (2025). PWM offline variable application based on UAV remote sensing 3D prescription map. *Artificial Intelligence in Agriculture*, *15*(3), 496–507. https://doi.org/10.1016/j.aiia.2025.01.011

Harun, R., Engku Ariff, E. E., Suhaimee, S., & Wan Mat, K. (2024). Farmers' perception towards the effective use of drones in selected rice farming areas in Malaysia. *International Journal of Agriculture, Forestry and Plantation*, *14*, 52–58.

Hunter III, J. E., Gannon, T. W., Richardson, R. J., Yelverton, F. H., & Leon, R. G. (2019). Coverage and drift potential associated with nozzle and speed selection for herbicide applications using an unmanned aerial sprayer. *Weed Technology*, *34*(2), 235–240. https://doi.org/10.1017/wet.2019.101

Iqbal, A., He, L., Ali, I., Ullah, S., Khan, A., Khan, A., Akhtar, K., Wei, S., Zhao, Q., Zhang, J., & Jiang, L. (2020). Manure combined with chemical fertilizer increases rice productivity by improving soil health, post-anthesis biomass yield, and nitrogen metabolism. *PLOS One*, *15*(10), e0238934. https://doi.org/10.1371/journal.pone.0238934

International Organization for Standardization. (2005). *Equipment for crop protection — Methods for field measurement of spray drift*. ISO. https://www.iso.org/standard/35161.html

Ismail, S. A., Yahya, A., Su, A. S. M., Asib, N., & Mustafah, A. M. (2021). Drone payload and flying speed effects on rotor blades' RPM and traveling pattern for agricultural chemical spraying. *Basrah Journal of Agricultural Sciences*, *34*(Special Issue 1), 157–170. https://doi.org/10.37077/25200860.2021.34.sp1.16

Jomantas, T., Lekavičienė, K., Steponavičius, D., Andriušis, A., Zaleckas, E., Zinkevičius, R., Popescu, C. V., Salceanu, C., Ignatavičius, J., & Kemzūraitė, A. (2023). The influence of newly developed spray drift reduction agents on drift mitigation by means of wind tunnel and field evaluation methods. *Agriculture*, *13*(2), 349. https://doi.org/10.3390/agriculture13020349

Lodwik, D., Pietrzyk, J., & Malesa, W. (2020). Analysis of volume distribution and evaluation of the spraying spectrum in terms of spraying quality. *Applied Sciences*, *10*(7), 2395. https://doi.org/10.3390/app10072395

Matthews, G. A., Bateman, R., & Miller, P. (Eds.) (2014). *Pesticide application methods* (4th ed.). Wiley-Blackwell.

Miller, P. C. H., & Ellis, M. C. B. (2000). Effects of formulation on spray nozzle performance for applications from ground-based boom sprayers. *Crop Protection*, *19*(8–10), 609–615. https://doi.org/10.1016/S0261-2194(00)00080-6

Montanarella, L., & Panagos, P. (2021). The relevance of sustainable soil management within the European Green Deal. *Land Use Policy*, *100*, 104950. https://doi.org/10.1016/j.landusepol.2020.104950

Nansen, C., Ferguson, J. C., Moore, J., Groves, L., Emery, R., Garel, N., & Hewitt, A. (2015). Optimizing pesticide spray coverage using a novel web and smartphone tool, SnapCard. *Agronomy for Sustainable Development*, *35*, 1075–1085. https://doi.org/10.1007/s13593-015-0309-y

Nhamo, L., Magidi, J., Nyamugama, A., Clulow, A. D., Sibanda, M., Chimonyo, V. G. P., & Mabhaudhi, T. (2020). Prospects of improving agricultural and water productivity through unmanned aerial vehicles. *Agriculture*, *10*(7), 256. https://doi.org/10.3390/agriculture10070256

Niemoeller, B., Harms, H. H., & Lang, T. (2011). Injection of liquids into the soil with a high-pressure jet. *Agricultural Engineering International: CIGR Journal*, *13*(2).

Nuyttens, D., Baetens, K., De Schampheleire, M., & Sonck, B. (2007). Effect of nozzle type, size and pressure on spray droplet characteristics. *Biosystems Engineering*, *97*(3), 333–345. https://doi.org/10.1016/j.biosystemseng.2007.03.001

Papademetriou, M. K. (2000). *Rice production in the Asia-Pacific region: Issues and perspectives*. Food and Agriculture Organization of the United Nations. https://fao.org/4/x6905e/x6905e04.htm

Pascuzzi, S., Cerruto, E., & Manetto, G. (2017). Foliar spray deposition in a "tendone" vineyard as affected by airflow rate, volume rate and vegetative development. *Crop Protection*, *91*, 34–48. https://doi.org/10.1016/j.cropro.2016.09.009

Rosedi, F. A., & Shamsi, S. M. (2022). Comparative efficacy of drone application in chemical spraying at paddy field MADA Kedah. In *IOP Conference Series: Earth and Environmental Science* (Vol. 1059, No. 1, p. 012002). IOP Publishing. https://doi.org/10.1088/1755-1315/1059/1/012002

Salyani, M., Zhu, H., Sweeb, R. D., & Pai, N. (2013). Assessment of spray distribution with water-sensitive paper. *Agricultural Engineering International: CIGR Journal*, *15*(2), 101–111.

Shan, C., Wang, G., Wang, H., Xie, Y., Wang, H., Wang, S., Chen, S., & Lan, Y. (2021). Effects of droplet size and spray volume parameters on droplet deposition of wheat herbicide application by using UAV. *International Journal of Agricultural and Biological Engineering*, *14*(1), 74–81. https://doi.org/10.25165/j.ijabe.20211401.6129

Siddharta, A. (2025). *Production of rice in Malaysia 2014–2023*. Statista. https://www.statista.com/statistics/794700/rice-production-volume-malaysia/

Srinivasarao, A., Khura, T. K., Parray, R. A., Kushwaha, H. L., & Mani, I. (2021). Spray droplet deposition, collection, and analysis techniques: A review. *The Pharma Innovation Journal*, *10*(6S), 156–165.

Taylor, W. A., Womac, A. R., Miller, P. C. H., & Taylor, B. P. (2004). *An attempt to relate drop size to drift risk*. ResearchGate. https://www.researchgate.net/publication/237287986_An_Attempt_to_Relate_Drop_Size_to_Drift_Risk

Tsouros, D. C., Bibi, S., & Sarigiannidis, P. G. (2019). A review on UAV-based applications for precision agriculture. *Information*, *10*(11), 349. https://doi.org/10.3390/info10110349

Wang, C., Wongsuk, S., Huang, Z., Yu, C., Han, L., Zhang, J., Sun, W., Zeng, A., & He, X. (2023). Comparison between drift test bench and other techniques in spray drift evaluation of an eight-rotor unmanned aerial spraying system: The influence of meteorological parameters and nozzle types. *Agronomy*, *13*(1), 270. https://doi.org/10.3390/agronomy13010270

Wang, P., Hanif, A. S., Yu, S.-H., Lee, C.-G., Kang, Y. H., Lee, D.-H., & Han, X. (2024). Development of an autonomous drone spraying control system based on the coefficient of variation of spray distribution. *Computers and Electronics in Agriculture*, *227*(Part 1), 109529. https://doi.org/10.1016/j.compag.2024.109529

Xue, X. Y., Tu, K., Qin, W. C., Lai, Y. B., & Zhang, H. H. (2014). Drift and deposition of ultra-low altitude and low volume application in paddy field. *International Journal of Agricultural and Biological Engineering*, *7*(4), 23–28. https://ijabe.org/index.php/ijabe/article/view/935/pdf

Youssef, M. M. A., & Eissa, M. F. M. (2014). Biofertilizers and their role in management of plant parasitic nematodes. A review. *E3 Journal of Biotechnology and Pharmaceutical Research*, *5*(1), 1–6.

Zhang, J., Xu, X., Lv, Y., Zhao, X., Song, J., Yu, P., Wang, X., & Zhao, E. (2024). Droplet deposition and efficacy of real-time variable-rate application of herbicides at reduced dose in winter wheat fields. *Agronomy*, *14*(1), 211. https://doi.org/10.3390/agronomy14010211

Zhu, H., Salyani, M., & Fox, R. D. (2011). A portable scanning system for evaluation of spray deposit distribution. *Computers and Electronics in Agriculture*, *76*(1), 38–43. https://doi.org/10.1016/j.compag.2011.01.003

Pertanika J. Sci. & Technol. 33 (6): 2685 - 2722 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Review Article

Time Series Prediction for Cycling Sport Performance Through a Machine Learning Predictor

Muhammad Zulazri Hanis Mohd Nawi[1]*, Sukhairi Sudin[1,2], Fathinul Syahir Ahmad Saad[1,2], Muhamad Khairul Ali Hassan[1,2], Faranadia Abdul Haris[3], Nurul Syahirah Khalid[1,2] and Kamarulzaman Kamarudin[1,2]

[1]*Faculty of Electrical Engineering and Technology, Universiti Malaysia Perlis, Pauh Putra Campus, 02600 Arau, Perlis, Malaysia*
[2]*Centre of Excellence for Advanced Sensor Technology, Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia*
[3]*School of Electrical Engineering, Universiti Teknologi Mara, Cawangan Pulau Pinang, 13500 Permatang Pauh, Pulau Pinang, Malaysia*

ABSTRACT

Predicting sports performance has become a central focus in sports analytics, driven by the increasing availability of data and the growing recognition of its potential impact on decision-making in the sports sector. Time series analysis and real-time prediction of athletic performance involve forecasting an athlete's performance over time, allowing coaches and sports scientists to refine training programs, manage workload, and make informed strategic decisions. This study thoroughly examines time series prediction and real-time prediction in sports, as well as the artificial intelligence (AI) techniques employed by prior researchers. The review is conducted with precision, strictly following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) criteria. This article examines time series prediction and real-time prediction methodologies that utilize machine learning (ML) and deep learning (DL) approaches, spanning the period from 2020 to 2025. This article covers the range of AI methodologies from the most basic to the most advanced models. A detailed assessment of ML and DL methodologies, grounded in prior research findings, is presented. All approaches examined in this paper significantly influence the primary future study, which focuses on the hybrid long

ARTICLE INFO

Article history:
Received: 01 January 2025
Accepted: 29 April 2025
Published: 14 October 2025

DOI: https://doi.org/10.47836/pjst.33.6.13

E-mail addresses:
zulazrihanis7298@gmail.com (Muhammad Zulazri Hanis Mohd Nawi)
sukhairi@unimap.edu.my (Sukhairi Sudin)
fathinul@unimap.edu.my (Fathinul Syahir Ahmad Saad)
khairulhassan@unimap.edu.my (Muhamad Khairul Ali Hassan)
faranadia@uitm.edu.my (Faranadia Abdul Haris)
syahirahkhalid@unimap.edu.my (Nurul Syahirah Khalid)
kamarulzaman@unimap.edu.my (Kamarulzaman Kamarudin)
* Corresponding author

© Universiti Putra Malaysia Press

short-term memory (LSTM) peephole integration with gated recurrent unit (GRU) for use in track cycling sports, the principal objective of the research. This research is consistent with the United Nations' Sustainable Development Goals.

Keywords: Cyclist performance prediction, hybrid predictor, performance classification, sports prediction, time series prediction

INTRODUCTION

Within the realm of athletics, there is a constant drive to further improve and strive for perfection. To achieve a competitive advantage, athletes, coaches, and analysts continually seek innovative approaches to enhance their performance. One example of this is the utilization of emerging technologies to acquire their skills more rapidly. The integration of AI is emerging as a transformative force in this dynamic landscape of sports performance enhancement. ML and DL are serving as the cornerstone of this transformation, with data-driven insights serving as the foundation. As a result of the application of technology in sports science to deliver real-time feedback, the way athletes receive input and improve their performance has undergone a revolutionary change in recent years. According to Vec et al. (2024), the exploitation of real-time data has revealed itself to be particularly beneficial when it comes to the provision of feedback to athletes.

The year 1997 marks the early application of AI in the sports industry. Prozone (Sports Business Journal [SBJ], 2023) was the first company to deploy AI in 1997 for the purpose of developing the first computer vision player tracking system in elite sports. This was the beginning of the field. A revolutionary computerized tracking system that can analyze movement patterns in a variety of sports was launched by Prozone (Di Salvo et al., 2006). Most of the company's efforts have been concentrated on quantifying not only the motion characteristics of professional football players but also the work rate ratios of these players during professional competitions. Not only does this method have the benefit of being applicable in official competitions, but it also has the potential to assist in the evaluation of high-level football performance.

According to SBJ (2023), the first officiating use case for Hawk-Eye computer monitoring in tennis was implemented in 2001. In addition to being known as the rapid reaction system, the Hawk-Eye system is a tool that helps referees quantify the accuracy of their penalties. Although the theory behind Hawk-Eye technology is quite straightforward, it is extremely accurate. It consists of eight to ten high-speed cameras, four computers, and a wide-screen display, and it determines whether tennis is played outside or indoors by analyzing the records of different ball positions. In some large-scale tennis courts, Hawk-Eye technology has evolved into an all-important tracking system. Tennis moves at a high speed during competitions; after landing, it is prone to generating issues and disputes about whether the tennis is in line or out of line, as well as other similar ones. Additionally, the

Hawk-Eye system can accurately record and judge ball landing point positions, thereby eliminating similar problems (L. Li & Shi, 2014).

One of the first known applications of AI in player personnel choices occurred in 2007, when Houston Astros consultant Ari Kaplan used it to make the case for Hunter Pence's promotion to the major leagues (SBJ, 2023). This was the first time AI had been used in this capacity. In 2009, AI was transformed into a new system that was given the name SportVU. This system is utilized in basketball sports and is applied through the ball. Additionally, a player monitoring camera system was presented to executives from the NBA. SportVU, developed by STATS, LLC, records data points for all ten players, the three referees, and the ball every thirty seconds. These data points measure speed, distance, player separation, and ball possession. According to an article written Richman (2018), the entire league would be equipped by the year 2013. It is planned to record, code, and categorize each movement, including every step, dribble, pass, shot, rebound, and other actions.

The year 2011 marked the year AI led the revolution in highlights through WSC Sports. The use of AI in this context involves the analysis of live sports broadcasts, which enables the identification of significant events in real time and the unlocking of limitless highlight combinations that can be easily distributed across many platforms. In the National Basketball Association (NBA), there are two stages of analysis. The first step is to conduct a real-time analysis of the game and determine what occurs during each second of the competition, from every goal to every touchdown. According to Lingeswaran (2023), each event is automatically detected and tagged with the precise day, time, and location at which it occurred. Additionally, they determine the beginning and ending of each play.

This evolution led Spiideo to introduce the first AI-powered, automated camera system for tracking sports in 2012. With the assistance of AI, it eliminates the friction associated with the roles of camera operator, producer, director, and commentator, thereby delivering production capabilities comparable to those of traditional broadcasting (Ruddock, 2024). A proliferation of chatbots from Satisfi Labs, Facebook Messenger, and Sapiens was observed in the sports industry in 2016. Satisfi Labs' first sports customer was the New York Mets (Anderson & Rainie, 2023). The Associated Press contracts Automated Insights to utilize its Wordsmith tool for game coverage on minor league baseball and high school sports that would otherwise go unreported (Mullin, 2016). Neither of these organizations would have been able to cover these types of sports otherwise.

Zone7 is the first sports science firm to use AI to anticipate the risk of athlete injuries. Their AI system can perform sophisticated computations on large amounts of data efficiently and consistently, thereby providing valuable insights daily. Due to its capability of performing these functions in a reliable manner in real-time, it reduces the amount of work required for data analysis. Additionally, it has the potential to make a significant contribution, even if it has not yet been fully utilized, to the way that professionals working

in multi-disciplinary teams (MDT) manage and minimize injury risks in professional athletes (Buchanan et al., 2022).

In 2019, the world was taken aback by the launch of a new technology that incorporates AI and is incorporated into the motion capture app, Physimax. This technology, which requires only a smartphone camera, plays a significant role in optimizing athletes' skills, techniques, and strategies by providing detailed feedback on motion data (Suo et al., 2024). A new regulation was implemented at the stadium, specifically for football matches, to reduce the risk of exposure to the coronavirus during the COVID-19 pandemic, which began at the end of 2019 and continued through 2020 and 2021. Express Access, powered by Wicket, will be introduced by the Browns in 2021 to provide fans with a more convenient and time-saving method for entering FirstEnergy Stadium. To provide participating fans with a touchless, faster, and more pleasant experience as they enter the stadium, Wicket's facial verification technology and computer vision analytics are utilized. According to Gribble (2021), this is the most recent initiative implemented as part of the Browns' Responsible Restart Plan. This plan was implemented in response to the COVID-19 outbreak. It will continue to be implemented to provide fans with the safest and most enjoyable experience possible when attending Browns games at FirstEnergy Stadium.

Semi-automated offside video-assisted review will be put through its paces at the 2021 Arab Cup, in preparation for the 2022 World Cup. During the Fédération Internationale de Football Association (FIFA) World Cup 2022, which begins on November 21 in Qatar, FIFA has announced that semi-automated offside technology will be used. This technology will serve as a support tool for both the video match officials and the on-field officials, assisting them in making offside decisions that are quicker, more accurate, and more reproducible on the most important stage of all (FIFA, 2022). There is also a partnership between WHOOP and OpenAI to produce WHOOP Coach, the first generative AI training assistant (WHOOP, 2023). In the meantime, International Business Machines Corporation (IBM) will introduce AI-generated video commentary for tournaments such as the Masters, Wimbledon, and the U.S. Open in 2023 (Hastings, 2023).

Sports Popular with AI

This historical occurrence, as documented above, demonstrates that AI technology can be incorporated into virtually every aspect of sports, including media, coaching, and player monitoring systems. This also includes almost every type of sport, such as cycling, football, tennis, baseball, or rugby, among others, that can be improved with the application of an AI system. This can be done to enhance aspects such as player performance, ball trajectory, or clear view in referee systems, such as the predominant system of positioning and division of responsibility used by football match officials worldwide. This section examines three popular sports that have implemented AI technology. The dynamic nature

of the sport itself, which necessitates consideration of various circumstances, contributed to the selection of this sport. Football, basketball, and cycling are the sports that will be examined in this article.

Several studies have already been conducted in the field of football or by earlier researchers, such as Aleza and Vetrithangam (2023), which apply AI to prevent errors in the refereeing process of a football match. This is because any error or poor decision made by the referee during a match has the potential to completely change the course of the match and result in heartbreaking defeats. As a result, the findings of their research center focus on examining ways to improve the decision-making process in football, while also highlighting the limitations of the process in cases that require interpretation. They concluded that AI could revolutionize the decision-making process not only in football but also in other sports. This is something that they discovered via their investigation. While this is happening, other researchers, such as H. Li et al. (2024), are implementing AI in football to study methods that may improve the quality of football instruction through AI and metaverse applications in mobile Internet environments. By combining the metaverse and internet or Internet of Things (IoT) applications, they successfully translated the development of football teaching quality into a research project on a 360-degree panoramic football teaching video distribution method. Additionally, researchers such as Gomes et al. (2013) are investigating the development of an Intelligent Decision Support System to predict the outcomes of football games, with a particular emphasis on betting procedures.

As is the case in football, playing basketball also involves coaching to identify areas where talented new players can be developed. It is therefore possible to improve the process of teaching and training sessions by utilizing the AI system. Additionally, the new player will only require a shorter amount of time to comprehend the material as compared to the conventional technique. An earlier researcher, Tingting (2015), has provided the collegiate basketball team with an Intelligent Teaching System based on the development of an AI system. Her teaching method is a multi-level, dynamic, interactive multimedia presentation system that possesses good interactivity and intelligence. She presents this system to the classroom. Teachers and students can be based on the requirements of teaching and learning, as well as the selection of effective learning tools, if a logical relationship is established between the information contained in multimedia. The practical test demonstrates that her research has the potential to increase the effectiveness of basketball instruction in physical education colleges and universities, and it also has a certain promotional value. Her study has accomplished the design goals and requirements that were anticipated. Other researchers, such as T. Xu and Tang (2021), have utilized an AI system in basketball to prevent sports injuries caused by collisions in the sport. The researchers conducted an in-depth investigation of the shooting action, as well as an analysis of the fundamental training activities and the sport recognition involved in basketball training. During this

interim period, Yan et al. (2023) have successfully implemented an AI system in basketball by introducing basketball shooting analysis. This is because it has the potential to help basketball players refine their shooting methods and accuracy, which in turn can enhance the efficiency of both games and training sessions.

Cycling, which is separated into numerous categories such as track cycling, road bicycle racing, cyclo-cross, mountain bike racing, freestyle bicycle motocross (BMX), and cycle speedway, is a different sport from the two sports mentioned above (Wood, 2020). Balance, coordination, and riding tactics are some of the fundamental skills that are needed for all cyclists to hone (Proulx & Smith, 2014). Although cycling sports are classified into a few categories, they all share a common core talent that is essential for all cyclists to develop. In more advanced fields, it may be necessary to possess extra skills, but the fundamental abilities remain the same. Using AI, Ahmadi et al. (2024) were able to accurately estimate radial and mediolateral force components while cycling by utilizing inexpensive sensor data and participant characteristics. This was made possible by research conducted by Ahmadi et al. (2024). The reason for this is that precise measurements of pedaling kinetics and kinematics are crucial for maximizing the effectiveness of rehabilitation, exercise training, and understanding musculoskeletal biomechanics. There is a strong correlation between the pedal reaction force, which is the primary external force in cycling, and the activity of the lower limb muscles, as well as the joint reaction forces. This force is essential for musculoskeletal models.

An alternative perspective is presented by researchers such as Wu et al. (2020) regarding the use of intelligent cycling equipment to establish personalized training intensities for individual cyclists. In the meantime, Ibrahim et al. (2017) reported that their team was utilizing AI to forecast the chance of bicycle near misses. This was done to avoid potentially hazardous situations that could result in severe consequences, particularly during competitions or races held in urban areas. Different from Karetnikov (2019), who used the application of AI into cycling to analyze for professional cyclists and propose a new framework that aims to provide the coach with a method with which he or she can predict the performance in an upcoming race based on recently completed training sessions and planned training sessions in the near future, this framework aims to provide the coach with a method that predicts the performance of the athlete in an upcoming race. In this manner, the coach can evaluate several planned alternatives to determine the most effective method of preparing the athlete for competition. Lastly, but certainly not least, N. Li et al. (2024) and their fellow researchers utilized AI to investigate whether it can accurately predict the maximum volume of oxygen (VO_2) that the body can utilize during exercise across various exercise intensities by combining muscle oxygen (MO_2) with heart rate (HR). They used twenty young athletes who had received extensive training to conduct the following tests: a ramp incremental exercise, three submaximal constant-intensity exercises, and three severe, exhaustive-intensity exercises.

The Affected Factors on Sports Player Performance

As a result of deploying the AI system described above, it is evident that numerous aspects influence the sports player. Whether it be cyclists, football players, or basketball players, all these factors have the same component. In this section, the factors that affected player performance will be explained, including physical, cardiovascular, psychological, and any other relevant elements. Athletic success and elite performance are frequently the product of the best possible fusion and coordination of all aspects of sports readiness (i.e., physical fitness, tactical and technical expertise, bioenergetic and neuromuscular capacities and abilities, anthropometric traits, biomechanics, physiological, emotional intelligence, creativity, or personality), which develop because of methodical, sustained sports preparation. In contrast, individual and team sports have different weights assigned to these attributes (Calleja-González et al., 2023). While certain individual sports demand a high degree of bioenergetic, neuromuscular, and other capacities, the technical and

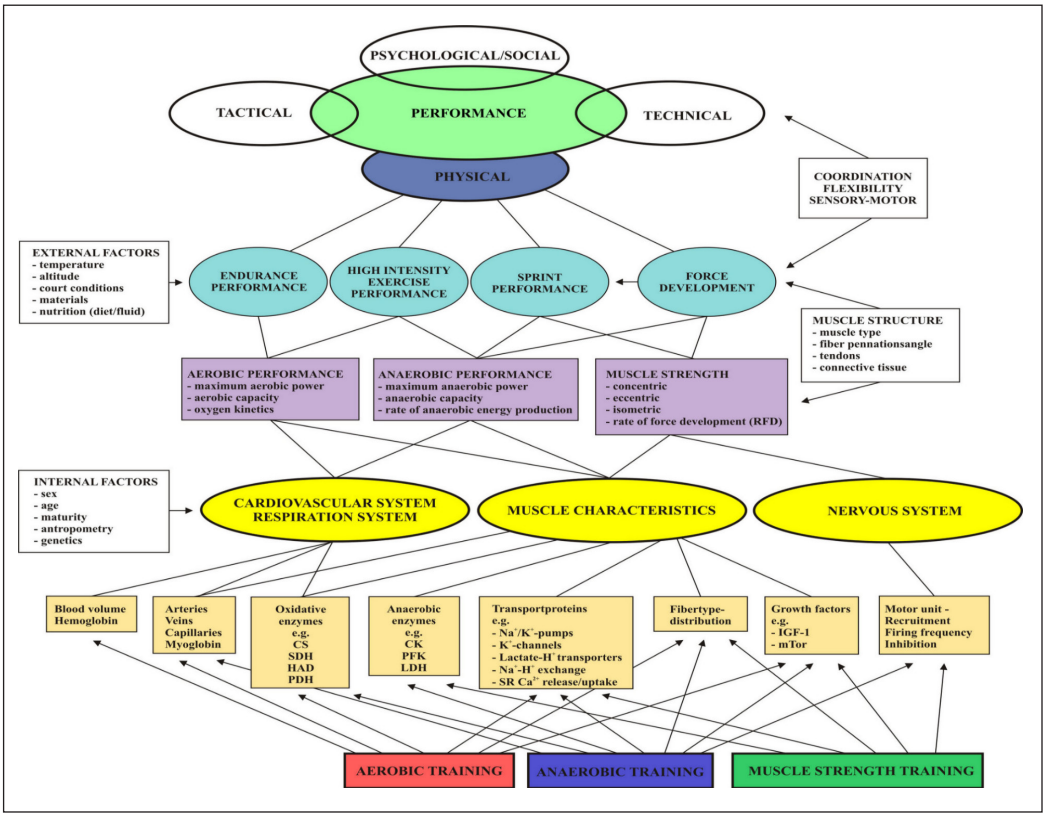


Figure 1. Comprehensive model of the factors that affect sports performance (Bangsbo, 2015)
Note. CS = Citrate synthase; SDH = Succinate dehydrogenase; HAD = Histamine dehydrogenase; PDH = Pyruvate dehydrogenase; CK = Creatine kinase; PFK = Phosphofructokinase; LDH = Lactate dehydrogenase; NA^+ = Sodium ion; K^+ = Potassium ion; H^+ = Hydrogen ion; SR = Sarcoplasmic reticulum; Ca^{2+} = Calcium ion; IGF-1 = Insulin-like growth factor 1; mTor = Mechanistic/mammalian target of rapamycin

tactical skills of team sports are intimately linked to performance and can make up for deficiencies in fitness levels (Bangsbo, 2015; Dolci et al., 2020). Bangsbo (2015) created a comprehensive model of the factors that affect sports performance in Figure 1.

Gavin (2022) asserted that a person's ability to succeed in sports is influenced by their sleep patterns, emotional condition (stress, anxiety, and sadness), pre-match routine, and romantic relationships. During his study on the factor of player performance, he did not take a bias towards any sport; instead, he exclusively focused on the factor.

According to the Global Web Index (GWI), as noted by Seedhouse (2017), since football is one of the world's major sports, various factors influence player performance based on their style of play. The factors encompass physical, psychological, tactical, and environmental aspects, all of which impact a football player's performance. Physical aspects include exhaustion and fitness levels. While physical fatigue (Dambroz et al., 2022) can negatively impact technical skills such as shooting, dribbling, and passing, high levels of both aerobic and anaerobic fitness are essential for maintaining performance throughout the game (Plakias & Michailidis, 2024). Players who possess higher levels of motivation and confidence tend to outperform others (Nakisa & Rahbardar, 2021), whereas those who experience high levels of stress and anxiety may struggle to make decisions and perform well overall. These psychological elements include motivation and worries (Nakisa & Rahbardar, 2021).

The two main tactical aspects are opposing quality and team strategy. A player's performance can be impacted by the quality of the opposition team, particularly in high-pressure situations, and the team's tactical approach, including formations and playing style, can affect individual performance (Bangsbo, 2014). The playing surface and weather conditions make up the environmental aspects. Extreme weather, such as heat or rain, can impair endurance and physical performance, and different types of playing surfaces, like grass or artificial turf, can affect performance and injury risk (Bangsbo, 2014). Coaches and athletes can enhance their training and match preparation to improve performance on the field by gaining a deeper understanding of these variables.

A basketball player's performance can be significantly influenced by several factors, including their level of physical fitness, psychological well-being, core training, performance analytics, shooting accuracy and technique, game circumstances, and external variables. It is essential to possess physical fitness qualities such as strength, speed, endurance, agility, and flexibility. A player's capacity to sprint, jump, balance, and change direction rapidly is influenced by these physical characteristics (S. Luo et al., 2023). Basketball players may be impacted by psychological issues since stress reduction, mental toughness, and attention are essential. Players who can remain composed under duress typically play better (S. Luo et al., 2023). Basketball players must engage in core training because total athletic success depends on having a strong core. Core training has been linked to improvements

in strength, agility, balance, passing, dribbling, shooting, and rebounding, according to studies (S. Luo et al., 2023).

Lastly, external circumstances and game conditions can significantly impact a player's performance throughout a match. Coaching techniques, crowd dynamics, and game location can all impact a player's performance. The comfort and confidence levels of a player might be impacted, for example, by playing at home vs. away (F. Wang & Zheng, 2022). The sum of these elements affects a player's overall performance on the court.

Between these sports, the factors that affect cyclists can be seen as having a significantly more profound impact on them if these factors occur. As for cyclists, the physiological, biomechanical, nutritional, aerodynamic, and physical components of elite cycling performance have all been examined (Atkinson et al., 2003; Craig & Norton, 2001; Faria et al., 2005a, 2005b). As further explanation, the individual cyclist factors, such as physiological attributes, such as aerobic capacity, muscle fiber composition, and lactate threshold (Phillips & Hopkins, 2020). Meanwhile, nutrition and hydration, as a proper diet and hydration, are crucial for maintaining energy levels and endurance (Phillips & Hopkins, 2020). Training and conditioning cannot be neglected, as these can contribute to individual cyclist factors that regular and targeted training can significantly enhance the performance of cyclists (Faria et al., 2005b).

To optimize a cyclist's performance during a match, tactical considerations are also necessary in cycling. Two examples of tactical variables in cycling sports are race strategy and interpersonal dynamics. Choosing the right pace, drafting position, and moment to put in your best effort all affect how well you perform; therefore, race strategy is crucial (Faria et al., 2005b). Teamwork, competitiveness, and interpersonal interactions with other cyclists are factors (Phillips & Hopkins, 2020). Wind, temperature, and humidity are just a few examples of the environmental elements that can have a significant impact on a player's performance during a match (Phillips & Hopkins, 2020). The effort required is also influenced by the kind of terrain, such as flat, hilly, or mountainous (Faria et al., 2005b).

Gear ratios, cadence, and bicycle layout are examples of equipment factors that can provide a unique match experience for every competitor. Bicycle setup, which encompasses both mechanical effectiveness and aerodynamics, is crucial in creating a distinct margin between players (Faria et al., 2005b). Pedaling cadence can increase efficiency throughout a match, and gear ratios and cadence can be used to adjust the ideal gear selection (Faria et al., 2005b). In addition, and just as importantly, socio-cultural and economic aspects of the world can influence the ease of resource acquisition, the strength of supportive networks, and the impact on possibilities for training and overall performance (Phillips & Hopkins, 2020).

To sum up, Lastella et al. (2015) stated that sleep and recovery are also one of the major patron preparations for cyclists, as cyclists with adequate sleep and recovery can

achieve better performance enhancement. It is possible for teams, coaches, and cyclists to create more successful training and competition plans when these aspects are understood.

The objective of this study is to investigate the current state-of-the-art in time series prediction and real-time prediction applications in sports, with a particular focus on cycling sports. However, this review will also encompass other sports, such as basketball and football/soccer, as well as research accomplishments, and it will recommend prospective future paths. A review of prior research on time series prediction in sports applications, primarily from the years 2020 to 2025, is presented in this publication. The purpose of this research was to examine the development of techniques over this time in chronological order. These techniques included ML and DL approaches, which were particularly concerned with real-time prediction as opposed to batch prediction when applied to datasets. This presentation is valuable since it is geared towards professionals in the sports industry, including sports scientists, coaches, and, most importantly, the players themselves, to motivate them to self-analyze overtime and maintain their performance in competition. The main contributions highlighted in this paper are:

- An in-depth investigation into how ML and DL models are applied in the realm of sports, as well as their respective methodologies.
- A methodical investigation of the application of time series prediction and real-time prediction, which is combined with ML and DL techniques and is utilized in the sports industry.
- The results of previous researchers on applying ML and DL in sports will be analyzed, and an evaluation of the field of applying these technologies in sports will be taken into consideration.
- Comparisons and discussions have been made regarding the most recent state-of-the-art procedures.

PROBLEM STATEMENTS

The primary objective of a more extensive body of study, which includes this review, is to utilize contemporary technologies to expedite the learning process in professional sports. It was estimated that the annual revenue of the professional sports sector is currently 159 billion dollars (Sim, 2024). This makes it a massive enterprise. This suggests that there is a significant possibility for any new technology that could provide the team using it with even the slightest advantage over its competitors. It is possible to acquire this competitive advantage by providing athletes with time series prediction feedback if they make a mistake, or by checking their performance while they are competing in sporting events (Hribernik & Kos, 2023). Currently, one of the most pressing questions on our minds is why the world needs AI, irrespective of whether it is ML or DL, in the sports business. In the sports business, human errors and faulty decision-making in football matches are two

of the issues that need to be addressed by AI (Aleza and Vetrithangam, 2023). A mistake like this could result in the referee making an incorrect decision during the match.

Additionally, according to the findings of other researchers, the teaching approach in physical education at universities continues to adhere to traditional teaching methodologies, particularly in the realm of sports, where concepts that are integral to the teaching work and should be enhanced accordingly are often overlooked (H. Li et al., 2024). The occurrence of sports injuries is an unavoidable issue for athletes participating in any sport, and it is also a circumstance that cannot be avoided during training and competition. The high-intensity and high-speed collisions that occur between players in basketball are one of the reasons why basketball is considered one of the sports that pose the greatest danger of injury. Because of this, we require the involvement of AI to either prevent it or, at the very least, mitigate the risk it poses to the player (T. Xu & Tang, 2021). In the meantime, the decision regarding the training plans that will lead to the greatest performance in the competitions is mostly based on the specialist knowledge of the team coach in the sport of cycling. Several recent advancements in the fields of sports engineering and data science have enabled the exploration of new avenues for data gathering and analysis driven by data. To support these decisions, it suggests new prospects for the use of AI, which is why the assistance from AI is required (Karetnikov, 2019).

However, before beginning to use AI, especially ML or DL, for event time series prediction to deliver feedback, the status of the subject must be investigated. The primary purpose of this study is to critically evaluate and examine works published between 2020 and 2025 that focus on the use of ML and DL algorithms to provide players with time series prediction and real-time feedback for any sport.

DATABASES, KEYWORDS, AND INCLUSION CRITERIA

In this paper, the PRISMA framework (Page et al., 2021) is followed for reviews. Three databases were searched, including Scopus and Web of Science to cover interdisciplinary research, as well as IEEE Xplore, which covers articles, conference proceedings, and other publications. The first database search was conducted in August 2024, during which relevant journals and conference proceedings were identified and extracted. To update the analysis, an additional search was conducted in October 2024 and February 2025 to include journal articles from December 2024 to February 2025. This paper aims to review all articles in the last six years that focus on using ML and DL to provide real-time feedback and time series prediction to athletes and to whatever sports they play, but mostly focused on cycling, and include other sports as well, like football and basketball, where they track the athletes using whatever equipment and sensors to capture. The following keyword string was used for Web of Science, IEEE Xplore, and Scopus: (“machine learning” OR “deep learning”) AND (real time OR “real-time” OR real-time OR concurrent OR instant) AND (sport OR

exercises) AND (time series OR prediction OR predictor OR predict) (time AND series AND prediction AND in AND sport).

Articles that did not provide a detailed description of the procedure were not taken into consideration. Additionally, publications that focused on the generalization of activities into major tasks, such as lying, sitting, or walking, were excluded from the compilation. Additionally, articles written in languages other than English were excluded. A summary of the inclusion and exclusion criteria is presented in Table 1. As a result of the fact that several of the papers satisfied more than one of the criteria for exclusion, the total is higher than the total number of publications that were not included.

Table 1
Inclusion and exclusion criteria

Inclusion	Exclusion
2020–2025	Outside 2020–2025
English	Full text not in English
The method is clearly described	No clear description
Active human movement	Not human movement
Time series prediction or real-time	Other than time series prediction or real-time
Specific sports activity	General activity

The scoping review primarily assessed real-time or time series prediction methodologies utilized in the current literature and investigated the accompanying research questions:

1. Data: What data sources were utilized? What was the magnitude of the sample size?
2. Preparation: Was dimensionality reduction implemented? What method was employed to partition the data set into training and validation subsets?
3. Methodology: What methodologies were employed? Do these methods consider the interdependencies among the segments or individual movements?
4. Validation: What measures were employed for performance assessment?

Study Selection

Keyword searches of the three databases retrieved 22,684 articles covering the period from August 2024, along with an additional 269 articles that included journal articles from December 2024 to February 2025. After screening the titles and abstracts, 3,250 articles remained for detailed full-text analysis. Of these 3,250 articles, 75 were relevant for inclusion. For quality assurance, each full text was screened, and data extraction was performed by using the Mendeley tool to extract the information systematically. An amount of 1,540 articles were not related to sports or addressed a different application area, 1,230 articles were excluded because the time series task was not included at all or was not one of the focuses of the article, 150 articles did not clearly describe the analysis or were not found, 125 articles that miss out from title and abstract screening and 110 articles were conference papers that were preliminary work for an included journal article that included the same methods or an extension of an existing work without any new method. Finally, 15 papers were excluded because their full texts were not in English. Figure 2 provides an overview of the inclusion process.

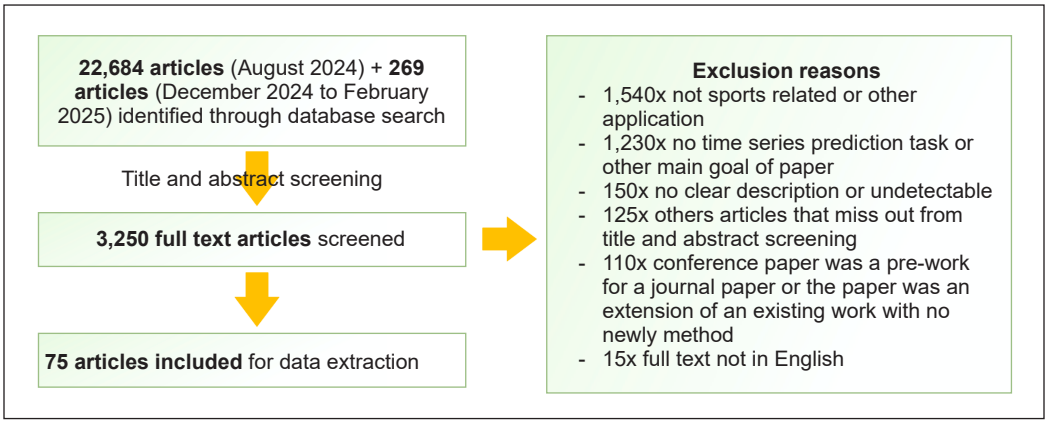


Figure 2. Flow diagram of the review

Recent Development of the Field

The field of using ML and DL in sports is developing rapidly. This paper splits the included articles by whether they use ML or DL. The reviewed articles indicate that performance analysis may be a vital tool to help coaches make decisions at the highest levels (Wright et al., 2013). There are many different prediction techniques, the two most common being linear and nonlinear (Zhang et al., 2022). Among these, nonlinear prediction should be used to forecast sports performance since a variety of factors influence sporting events. The physical and psychological state of the participants, the strategies used in the game, biomechanical, and even environmental aspects might have an impact on the outcome (Elmagd, 2019). As a result, although the nonlinear approach may take these elements into account, the linear prediction technique cannot work the same.

Nowadays, to improve decision-making and obtain insights into sports performance, sports organizations are depending more on advanced prediction models (Cojocariu, 2022). Traditional analytical techniques have given way to more dynamic, data-driven methods, which allow for a complete understanding of the complex aspects affecting sports results (Fujii, 2021). When working with time series data, data analytics becomes very important in the sports industry. The forecasting of time series athletic performance has been a crucial focus, realizing the necessity of identifying patterns and trends in sequential data (Kolambe & Arora, 2024). There is little doubt that predicting performance is difficult, and predicting long-term performance is much more difficult, with little research done on the subject to date (Pantzalis & Tjortjis, 2020).

Hence, below we split the application of time series prediction into two categories: time series prediction and real-time prediction using ML techniques, and time series prediction and real-time prediction using DL techniques. The taxonomy of ML and DL, as shown in Figures 3 and 4, helps to distinguish between techniques in ML and those in DL.

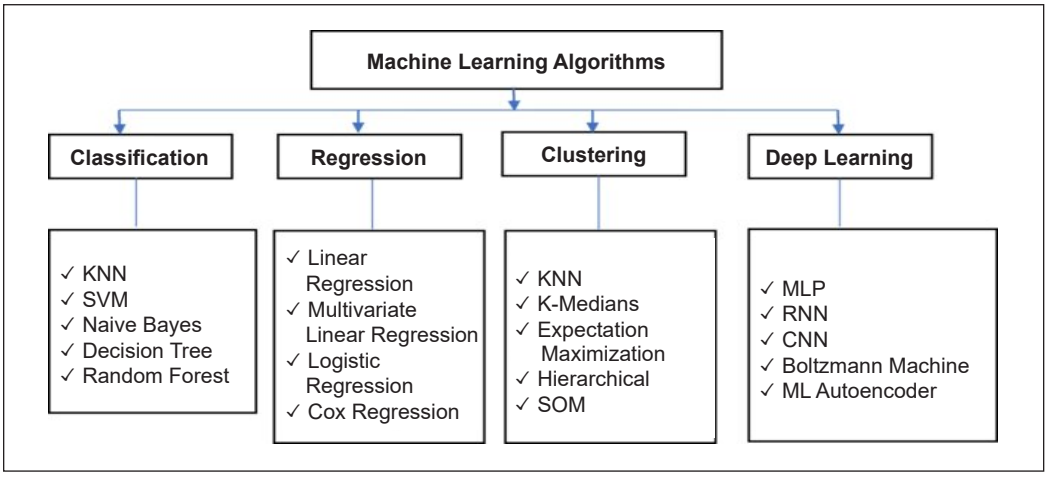


Figure 3. Taxonomy of machine learning algorithms (G. Wang et al., 2020)
Note. KNN = k-nearest neighbors; SVM = Support vector machine; SOM = Self-organizing map; MLP = Multilayer perceptron; RNN = Recurrent neural network; CNN = Convolutional neural network; ML = Machine learning

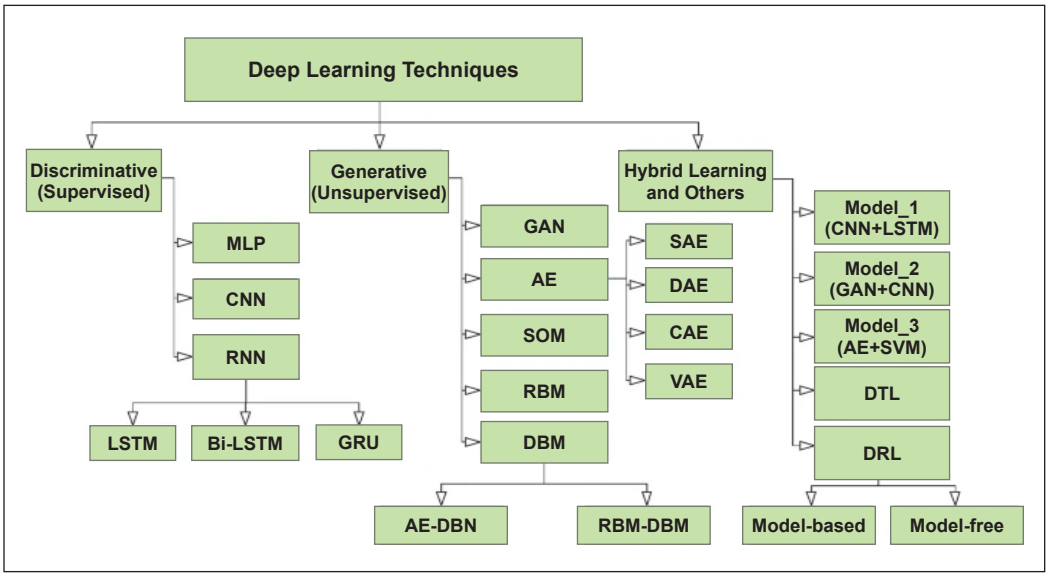


Figure 4. A taxonomy of deep learning techniques by Sarker (2021) was broadly divided into three major categories: (i) deep networks for supervised or discriminative learning, (ii) deep networks for unsupervised or generative learning, and (iii) deep networks for hybrid learning and others
Note. MLP = Multilayer perceptron; CNN = Convolutional neural network; RNN = Recurrent neural network; LSTM = Long short-term memory; Bi-LSTM = Bidirectional long short-term memory; GRU = Gated recurrent unit; GAN = Generative adversarial network; AE = Autoencoder; SAE = Sparse autoencoders; DAE = Denoising autoencoders; CAE = Computer-aided engineering; VAE = Variational autoencoders; SOM = Self-organizing map; RBM = Restricted Boltzmann machine; DBN = Deep belief network; AE-DBN = Autoencoder-Deep belief network; RBM-DBN = Restricted Boltzmann machine-Deep belief network; DTL = Decision tree learning; DRL = Deep reinforcement learning

ML METHOD OF TIME SERIES AND REAL-TIME PREDICTION

Linear Regression

Linear regression analysis is used to forecast the value of one variable depending on the value of another variable. The variable that the user intends to forecast is referred to as the dependent variable. The variable employed by the user to forecast the value of another variable is referred to as the independent variable (IBM, 2021). This technique was employed by the researcher (Yuan, 2024) to forecast the outcomes of certain basketball athletes involved in NBA competitions. The outcome of employing this technique indicates that linear regression emerges as the most effective model among the other models they tested for forecasting the performance of specific basketball players. The outstanding performance of the linear regression model, evidenced by its minimum mean squared error (MSE) of 0.0, implies an impeccable fit to the training data. Other researchers who have utilized linear regression analysis in their studies include De Beéck et al. (2018). Their research aims to forecast the rating of perceived exertion (RPE), a validated subjective metric of fatigue, utilizing inertial sensor data from individuals running outside. They attained a mean absolute error (MAE) ranging from 1.89 to 2.02.

Multiple Linear Regression

There was a study in 2020 that employed multiple linear regression (another variant of linear regression method) to forecast football player performance with the dataset consisted of 59 central defenders having played at least 10 matches for the English Premier League 2016–17 season and the result they obtained was 0.907 R^2 and 0.88 adjusted R^2 with the thirteen features were proved to be statistically significant (Pantzalis & Tjortjis, 2020).

Least Absolute Shrinkage and Selection Operator (LASSO) Regression Model

Another form of this structure (linear regression) is termed the LASSO regression model, which incorporates a penalty term to deter the coefficients of the independent variables from becoming excessively large (Wohlwend, 2023). The researchers who employed this technique are Shukla et al. (2022), which is utilized to forecast the overall revenue earned by a sports facility. The lasso regression model provides the most accurate forecast of total income earned by any sports facility on a specific day of the month, with an R^2 score of 0.97965, a MAE of 63.66423, and a root mean square error (RMSE) of 82.02202.

Generalized Linear Model (GLM)

Aarons et al. (2023) employed the GLM technique, a versatile extension of ordinary linear regression (Pennsylvania State University [PSU], n.d.), to predict the end-of-match outcomes of Australian football matches in real-time. Their classification accuracy was

initially low at the start of the match (45.7–48.8%) but improved significantly, reaching a peak near the match's conclusion (87.2–92.7%).

Support Vector Machine (SVM)

SVM is a supervised ML technique employed for classification and regression tasks. Although capable of addressing regression issues, SVM is especially adept at classification tasks (GeeksforGeeks, 2025d). Cui et al. (2023), demonstrated an application of this technique in wearable devices to analyze data related to body recovery and injury prevention in physical education. This approach facilitates the real-time collection of students' exercise data, encompassing metrics such as exercise volume, heart rate, steps, and distance using wearable technology. Their functional movement screening (FMS) test results ranged from 0.0756 to 0.0838, depending on the assessed activity.

From another perspective, Zheng and Yuan (2022) employed the SVM motion analysis model technique to analyze the time series data about the prediction and features of dance sport movements, achieving a recognition accuracy of 97 to 98%. Besides that, Hu (2024) utilized it for sports performance prediction through time series analysis of the physical performance of college students. He asserts that this prediction holds significant reference value for the university sports management department in planning and developing physical education courses, aligning the curriculum more closely with the actual needs of students and thereby establishing a more rational and scientific framework for college student physical training and daily exercise. The outcome of his proposed method, determined by the relative inaccuracy at each stage, ranged from 1.56 to a minimum of 0.06%.

Ren et al. (2024) also employed this strategy in their study on a real-time injury monitoring system. The motion depth learning technique was employed in their paper to identify activities of human motion targets. A real-time monitoring system for sports injuries was developed, encompassing the monitoring of human physiological data and the detection of target actions in human motion. The outcome achieved was 94.2% precision and 92.5% sensitivity.

Gradient Boosting

Gradient boosting is a widely utilized boosting approach in ML for classification and regression tasks. Boosting is a type of ensemble learning strategy that trains models progressively, with each subsequent model aiming to rectify the errors of its predecessor (GeeksforGeeks, 2025c). It amalgamates multiple weak learners to form robust learners. Yuan (2024) employed this technique during testing with linear regression, yielding a relatively low MSE of 0.05 for the gradient boosting model. Although it does not match the accuracy of linear regression, it still exhibits a significant level of precision in forecasting player performance in basketball.

eXtreme Gradient Boosting (XGBoost)

XGBoost is a scalable and distributed form of the gradient boosting framework, specifically a gradient-boosted decision tree (GBDT) ML library (NVIDIA Corporation, n.d.). Miah et al. (2022) employed this strategy in their study to anticipate and provide recommendations for both small and large sports teams regarding the suitability of an athlete for a specific game, utilizing a mobile application and mobile health data (m-health). The m-health utilizes mobile devices to collect clinical health information and monitor patients' vital signs in real-time. XGBoost outperformed the other studied ML algorithms, achieving an accuracy of 95.2%, a sensitivity of 99.5%, a specificity of 99.5%, and an F1 score of 99.66%. Zhu (2024) also employed the XGBoost approach to concentrate on real-time predictions of NBA game results. According to his findings, XGBoost models attained the maximum accuracy, roughly 99.88%. These results demonstrate their enhanced efficacy in forecasting NBA game results. He asserts that this will provide substantial promise for the integration of high-dimensional statistics and DL in sports analytics.

k-nearest Neighbors (KNN)

KNN is a non-parametric algorithm, a supervised learning classifier that uses proximity to classify or predict the categorization of a certain data point. It is a widely used and straightforward classifier for classification and regression in contemporary ML (IBM, n.d.b). Due to the enthusiasm surrounding the topic, Amendolara et al. (2023) conducted a comprehensive review of prior studies that employed this strategy to predict sports injuries. Their study states that KNN has practical restrictions regarding the sample sizes it can efficiently analyze. Nonetheless, its simplicity and versatility are evident. The incorporation of specialized sensors for enhanced data collection has refined KNN injury identification models and augmented their capacity to discern elements contributing to damage. The improved identification of predictive injury characteristics at the individual athlete level enables coaches and medical staff to modify training techniques to mitigate the recognized risk of injury.

Bayesian Inference

Bayesian analysis, a statistical inference method named after English mathematician Thomas Bayes, enables the integration of previous knowledge of a population parameter with evidence derived from sample data to inform the statistical inference process. A prior probability distribution for the parameter of interest is established initially. The evidence is subsequently acquired and integrated using Bayes' theorem to yield a posterior probability distribution for the parameter. The posterior distribution serves as the foundation for statistical judgments regarding the parameter (Britannica, 2025). In short, Bayesian inference is a method of statistical inference wherein the statistician gives subjective probability to the potential distributions that may produce the data (Taboga, 2021).

Autoregressive Integrated Moving Average (ARIMA)

ARIMA is a distinct methodology for forecasting time series data. ARIMA synthesizes these two methodologies, therefore its designation. Forecasting is a domain of ML that utilizes the historical behavior of a time series to predict one or more future values of that series (Noble, 2024). Almaleck et al. (2024) optimized this technique in their research to predict the 24-hour-ahead electrical usage of sports venues. The primary experimental implementation was conducted at Odd Football Club's Skagerrak Arena in Norway, yielding results of 15.29% for mean absolute percentage error (MAPE) and 30.61% for MAE, both of which are somewhat elevated. However, upon integrating the basic ensemble method (BEM) of neural networks with ARIMA to formulate a hybrid model, the value alters significantly to 9.03% and 20.24%. While the research conducted by his team appears to diverge from others by focusing on the electricity consumption of sports events rather than predicting sports performance, the methodologies employed by his team may be evaluated and utilized in our future research endeavors.

Random Forest

Random forest is a widely utilized ML technique, which amalgamates the outputs of numerous decision trees to produce a singular outcome. The simplicity and adaptability of the system have driven its widespread use, as it effectively addresses both classification and regression challenges (IBM, n.d.a). Stübinger et al. (2020) employed this technique in their research, comparing various models to predict the outcomes of football matches based on match and player attributes. Among all the models evaluated by their team, the random forest technique had the highest accuracy of 81.26%, the lowest RMSE, and the least mean absolute deviation (MAD).

Others

Other researchers created their own model of AI or used methods that were not specified. The researchers, like N. Li et al. (2024), developed an ML model to predict volume of oxygen (VO_2) at various exercise intensities by integrating muscle oxygen (MO_2) and HR in cycling. Their findings indicated that all models yielded comparable results, although the accuracy of VO_2 predictions varied among the models at different exercise intensities. Notably, the left model (LM) + right model (RM) + heart rate (HR) model exhibited a robust correlation, producing an r value of 0.94 with a significance level of $p < 0.001$.

Other researchers, Eid et al. (2024), also developed an ML hybrid model that illustrates the effectiveness of the Sport AI Model (SAIM) framework in improving the precision of sports-related outcome predictions. They evaluated their models to analyze English Football League sports data, aiming to gain in-depth insights into player performance in the English Football League. They assert that the incorporation of big data analytics, facilitated

by the SAIM model, can transform football strategies, enhance player performance, and promote data-driven decision-making, noting that their model has successfully predicted player performance in real-time.

There is also a unique method that was applied in their research from Almarashi et al. (2024), who employed a distinctive methodology in their research, utilizing a dynamic technique that examines performance structure through both linear and nonlinear time series models, specifically the non-linear autoregressive neural network (ARNN) model, alongside traditional stochastic linear and nonlinear models such as ARIMA, Exponential Smoothing (ETS), and Trigonometric Seasonal Box-Cox autoregressive moving average (ARMA) Time Series (TBATS). Their research focuses on developing predictive models for performance metrics that can aid both tennis players and sports analysts in predicting player rankings in upcoming matches. Their findings indicate that the ARNN model surpasses all other competing models, as seen by lower values of RMSE, MAE, and MAPE. The prediction results derived from the ARNN model indicate low 95% confidence intervals, signifying enhanced accuracy and dependability in the forecasts.

Table 2 presents the best-performing methods from across all methods used in ML to predict sports-related activities, incorporating time series prediction, real-time prediction, or capturing temporal dependencies.

Summary

The entire body of research publications and journals that dealt with the application of ML techniques to the prediction of time series or real-time predictions for sports applications was examined. The period covered by the paper that was examined is between 2020 and 2025. A thorough examination of the application in sports, the performance of the AI models, and the outcomes that they achieved were the primary considerations that led to the selection of the findings from the comparison table. Each of the findings from the comparison table will also serve as a guide for the subsequent study, which will be conducted in this article, focusing on the construction of the Peephole-GRU Predictor for Cyclist Player performance prediction. A further observation that can be made is that between the years 2020 and 2025, there was a minimal number of researchers who conducted time series prediction using the ML method. Furthermore, a small number of researchers successfully validated the AI model in practical situations, particularly in sports applications. Study by N. Li et al. (2024) became one of the main references (apart from the DL method) for our research. This is because their method utilizes the LM + RM + HR model, which they have created and modified from the vanilla ML method, to predict VO_2 across different exercise intensities by combining MO_2 with HR in cycling sports. This method was chosen from among the top five models in the comparison table. The reason for this is not only that the model they have adapted is highly suited and compatible

Table 2
Comparison of the best methods (best of five) by previous researchers based on results obtained for implementing time series prediction and real-time prediction in machine learning methods

References	Algorithm method	Result			
		Model concept	Accuracy	Limitation	Application
Zheng and Yuan (2022)	SVM	Vanilla	97–98% for recognition accuracy	Since the structure used does not have a higher level of supervision, the effect is not ideal in some aspects	Prediction of characteristics of dance sport movements
Aarons et al. (2023)	GLM	Upgrade the variant of linear regression	Classification accuracy was low at the start of a match (45.7–48.8%) but increased to a peak near the end of a match (87.2–92.7%)	Only considers technical performance through technical PIs; other aspects of performance are not accounted for. Also, the predictions are based on the performances of all teams in the AFL, which do not provide insight into the specific game styles of individual teams	Predicting the end-of-match outcome of Australian football matches in real-time
Zhu (2024)	XGBoost	Upgrade the variant of gradient boosting	Achieved the highest accuracy, reaching approximately 99.88% compared to another model	Required substantial computational resources, challenges in interpretability, and necessitating additional tools to understand the decision-making process; also, insufficient or noisy data could hinder the model's ability to learn meaningful patterns	Real-time prediction of NBA game outcomes
N. Li et al. (2024)	LM + RM + HR model	Unspecified ML Model	Very strong correlation that produce ($r = 0.94$, $p < 0.001$)	Complexity and time-consuming to train and implement the model	To predict VO_2 across different exercise intensities by combining MO_2 with HR in cycling sport
Almarashi et al. (2024)	ARNN	Hybrid	95% confidence intervals, indicating higher accuracy	Complexity and time-consuming to train and implement the model	Predicting player performance in tennis

Note. SVM = Support vector machine; GLM = Generalized linear model; PIs = Performance indicators; AFL = Australian Football League; NBA = National Basketball Association; XGBoost = eXtreme Gradient Boosting; LM = Left model; RM = Right model; HR = Heart rate; ML = Machine learning; VO_2 = Volume of oxygen; MO_2 = Muscle oxygen; ARNN = Autoregressive neural network

with their parameters, but also that the application involving cycling sport is the same as the one that will be conducted in the future. Furthermore, the data that they gave, which demonstrated a substantial correlation ($r = 0.94$, $p < 0.001$), further enhances our confidence in their empirical research.

DL METHOD OF TIME SERIES AND REAL-TIME PREDICTION

Artificial Neural Network (ANN)

ANN is a technique in AI that enables computers to interpret data in a manner analogous to the human brain (Amazon Web Services [AWS], n.d.). Ahmadi et al. (2024) demonstrated the application of this technique to estimate radial and mediolateral forces, offering a cost-effective solution for studying pedaling biomechanics using stationary cycle ergometers. This research revealed that a DL model can precisely forecast radial and mediolateral force components during cycling by utilizing cost-effective sensor data and participant attributes. The neural network model achieved an inter-subject normalized root mean square error (nRMSE) of 0.15 ± 0.02 and 0.26 ± 0.05 for radial and mediolateral forces, respectively, at high cadence, and 0.20 ± 0.04 and 0.22 ± 0.04 at self-selected cadence. The model exhibited diminished accuracy in forecasting mediolateral forces compared to radial forces, which could be enhanced by increasing the sample size and exploring a wider range of power outputs.

Additionally, Stetter et al. (2020), who utilized it to predict knee joint forces (KJF) during athletic activities, based on data acquired from wearable sensors. Thirteen individuals were outfitted with two inertial measurement units (IMUs) positioned on the right leg. Participants executed a range of actions, encompassing linear motions, directional changes, and jumps. Biomechanical modelling was conducted to ascertain KJF. Their model was developed to represent the correlation between the IMU signals and the KJF time series. The results demonstrated that the estimation accuracy of the ANN varied among movements, although it remained satisfactory for most of them.

Radial Basis Function Neural Network (RBFNN)

RBFNN, another variant of ANN, employs radial basis functions as activation functions. The RBFNN is a three-layered feedforward neural network. The initial layer is linear and solely disseminates the input signal, but the subsequent layer is nonlinear and employs Gaussian functions. The third layer linearly aggregates the Gaussian outputs. During training, only the weights connecting the hidden layer to the output layer are adjusted (Alireza, n.d.). Oytun et al. (2020) applied this method to predict specific types of athletic performance in female handball players, involving countermovement jumps with hands-free and hands-on-hips, 10-m and 20-m sprints, a 20-m shuttle run test, and a handball-specific

agility test. The study's results indicated that the radial-basis function neural network effectively predicted the examined forms of athletic performance, with R^2 scores ranging from 0.86 to 0.97.

Deep Neural Networks (DNN)

DNN is a type of ANN that has many deep layers, meaning there are many layers in between the input and output layers. Deep architecture facilitates the learning of more intricate models than shallow architecture. However, they also increase the quantity of significant training parameters, including the number of layers and the number of units per layer (Salvi et al., 2021). Zhao and Li (2023) proposed a study utilizing DNN to predict the risk of sports injuries in athletes following after-school physical exercise, which encompasses seven movements: squat, hurdle step, straight leg squat, shoulder flexibility, active straight leg lift, trunk stability push-up, and body rotation, involving 29 athletes. The validated DNN model achieved an accuracy of 88.10%.

J. Xu (2022) used the DNN method to analyze predictions of sports competition results, as accurate predictions enable educators to implement suitable training for students, hence optimizing learning outcomes. The model they tested has a significant impact on performance prediction, and the values generated by their prediction model can minimize systematic prediction errors, thereby enhancing the accuracy of performance forecasts for sports competitions in China, their primary focus. The extracted values, with an absolute prediction error of 1.6167 and an RMSE of 0.142, substantiated the efficacy of their AI model as claimed.

Zhou (2022) employed this strategy to forecast a player's performance accurately. Precise forecasting of athletic performance is beneficial not only for athletes but also aids in the advancement of sports. They assert that employing this approach has enhanced the accuracy of sports performance predictions and reduced prediction error in comparison to conventional methods. The prediction error value obtained with their AI model is 0.12, in contrast to the traditional method's value of 0.15.

Generative Adversarial Networks (GANs)

GANs are one of the types of ANN algorithms that are a category of neural networks that independently discern patterns within input data to produce novel instances that resemble the original dataset (GeeksforGeeks, 2025b). G. Li (2022) employed this method to assess the prediction of sports training performance based on physical fitness data of college students. The outcome achieved from their proposed model during the testing phase is 98% accuracy at the conclusion of the forecast. This demonstrated that their model attains superior categorization prediction outcomes compared to the other models they tested.

Convolutional Neural Network (CNN)

CNNs are a distinct kind of neural network specifically engineered to analyze grid-structured input, including images. They are especially adept at identifying and processing pictures. Inspired by the visual processing mechanisms of the human brain, CNNs proficiently capture hierarchical patterns and spatial dependencies in images (GeeksforGeeks, 2025a).

In 2024, G. Wang and Ren employed this strategy to precisely forecast college students' athletic performance and incentivize them to enhance their physical condition. The conclusion of their studies indicated an accuracy of 0.93, a precision of 0.94, and an F1-score of 0.93. Their test demonstrates the suggested model's superior predictive ability, validating its effectiveness in accurately forecasting sports achievements in accordance with actual criteria. From different perspectives, Pappalardo et al. (2020) offer a unique methodology for training a time series classifier utilizing a CNN to predict the likelihood of a player sustaining an injury in a future time frame, based on their historical workload in football. The results they achieved were 0.92 in precision and 0.88 in F1-score. This leads them to assert that their trials, conducted using injury records and workload data, validate the efficacy of the prediction model.

Imran (2022) suggested a CNN method for recognizing sporting activities to be utilized within wearable sensors. Wearable sensors can generate time-series data from accelerometers, gyroscopes, and magnetometers. Their AI model comprises 1,251 trainable parameters, achieving a test accuracy of 98.387% across six activities: badminton, basketball, cycling, football, skipping, and table tennis. The average F1 score, recall, and precision are 98.0, 98.7, and 98.0%, respectively.

LSTM

LSTM is an advanced variant of the recurrent neural network (RNN) that effectively captures long-term dependencies in sequential data, rendering them suitable for tasks such as language translation, speech recognition, and time series forecasting. In contrast to conventional RNNs that utilize a singular hidden state propagated through time, LSTMs incorporate a memory cell that retains information over prolonged durations, effectively tackling the issue of learning long-term dependencies (GeeksforGeeks, 2025e).

Nokihara et al. (2023) employed this methodology to analyze data, enabling players to gain an advantage in the rapid rally progression of badminton matches. They explored the innovative task of forecasting future shuttle trajectories in badminton match videos by detecting players and training a time-sequence model. They confirmed that the LSTM employed in this approach achieves maximum accuracy by transitioning to a time-sequence model. Two sorts of displacement errors were utilized as evaluation measures for the experiment. The first metric is the average displacement error (ADE), defined as the mean of the errors across all output frames. The second is the ultimate displacement error

(UDE). The mean value obtained for the ADE is 0.04908, whereas for the fractional-order differential equation (FODE), it is 0.08391.

Sadr et al. (2025) employed an LSTM technique to identify and predict sports injuries for the management of athletes' performance and health. The results obtained demonstrate that the LSTM model attained the best accuracy of 91.5%, confirming its efficacy in identifying and forecasting injury patterns in athletes. Their findings emphasize the potential for incorporating DL methodologies into sports injury management systems to improve early identification and preventive efforts.

The LSTM method was also employed by Hilmkil et al. (2018) to predict the heart rate response of a cyclist during a training session. Their LSTM model was trained for roughly one week on a Titan X Pascal GPU, achieving final minimum, mean, and maximum RMSE values of 2.51, 5.62, and 25.67, respectively, on the validation set. This demonstrated that it is feasible to train a model capable of predicting a cyclist's heart rate at any specific moment.

A different view of the idea taken by L. Wang et al. (2021) involves the integration of knowledge about big data-driven computer vision and DL algorithms, specifically LSTM, to facilitate intelligent labelling and representation of individual human movements depicted in video sequences for gymnastics. The results obtained from the LSTM model demonstrate a 14.22% improvement over the other models tested for sports video categorization and prediction.

Optimal athletic performance necessitates a balance between rigorous training and sufficient recovery. Because of this, Biró et al. (2024) employed LSTM to develop a resilient AI model aimed at forecasting weariness and stamina. Nineteen regular, injury-free runners from University College Dublin, Ireland, were used as a study dataset in the experiment, under both tiredness and non-fatigue conditions. The optimal outcome achieved from their evaluation of this model is a precision of 0.60, indicating that 60% of the predictions made were accurate. The recall and F1 score are both 0.60, reflecting precision and recall, which suggests a comparably balanced performance with a slight preference for accuracy. The results indicate that their LSTM model has reduced accuracy in forecasting fatigue levels, but they attribute this to observed performance improvements. The proposed enhancement methods may encompass additional hyperparameter optimization (modifying the quantity of LSTM units, altering the learning rate, or testing various activation functions), reassessing the feature engineering approach, or utilizing strategies to mitigate overfitting, which could be a concern due to the extensive number of training epochs in their research. The excessive number of epochs in the training process may lead to issues with computing efficiency and an increased potential for overfitting, especially if validation accuracy stabilizes early in the training phase. Consequently, it is essential to assess the model's performance on the validation set during the training phase to optimize the number of epochs and avert overfitting.

Others

Other researchers created their own models of AI or used hybrid methods or unspecified methods. One of the hybrid modifications of LSTM that is an outstanding idea comes from Zhang et al. (2022), which integrates the LSTM model with an attention mechanism and a time sliding window, resulting in the development of an attention-based sports-aware LSTM (AS-LSTM) designed to forecast outcomes in football matches. The results derived from their comparative table of actual and anticipated outcomes indicate that the model's accuracy reached 80% during the performance evaluation of the AS-LSTM model prediction. Consequently, they assert that their AS-LSTM model is accurate and helpful in predicting football match outcomes based on teams' past performance in competitions.

Watson et al. (2021) implemented an innovative hybrid concept model that utilized multi-input RNN and CNN to forecast the outcomes of sequences in Rugby Union, based on the ordered sequence of actions and their locations on the pitch. They assert that the CNN-RNN was the most effective single-input model across six outcomes when validated in real-world applications.

The soft actor–critic (SAC)-optimized DL model, developed by Tang et al. (2025), aims to enhance real-time monitoring and analysis of track and field athletes, overcoming the shortcomings of conventional monitoring systems regarding real-time performance and precision. The system was constructed based on a deep reinforcement learning algorithm. They have asserted that their study validates the considerable benefits of their technology over conventional approaches in terms of response speed, accuracy, and energy efficiency. Their proposed solution, the SAC-optimized system, enhances response time (200 ms compared to 250 ms) and data processing accuracy (98.5%) in real-time monitoring, especially for intricate sports such as high jump and discus throw.

Chen and Yuan (2021) applied a CNN algorithm hybrid with a residual network approach in their research on a sports injury prediction system utilizing visual analysis technology. They assert that their model has achieved one of the best accuracies on the training set, as evidenced by a final loss of only 0.6 at the conclusion of the forecast. This result demonstrates that their recommended strategy is effective.

Factorization deep product neural network (FDPNN) is a hybrid DL model aimed at enhancing predictive performance through the integration of factorization methods and DL (He & Chua, 2017). The primary components included are factorization machines (FM), product layer, and DNN (Guo et al., 2017). The researcher who incorporated this hybrid into their study was ZhaoriGetu (2022), to predict pupils' sports performance in school. They are evaluating various quantities of hidden layer neurons, with 256 emerging as the optimal choice, resulting in the FDPNN model achieving superior predictive performance in sports compared to alternative methods. The obtained values for accuracy, precision, and F1 score were 0.8668, 0.9499, and 0.9071, respectively.

R. Luo and Krishnamurthy (2023) introduces an innovative hybrid DL approach, termed GATv2-TCN, for forecasting player performance in basketball. This model is a hybrid AI framework that integrates a graph attention network with a temporal convolution layer, surpassing prior models that depend exclusively on player historical performance. The performance comparison reveals that the RMSE and MAE values for their model are 2.222 and 1.642, respectively, in contrast to other models examined in their study, which produced significantly higher values. They assert that this demonstrated its efficacy compared to other models they evaluated and is already yielding money in sports betting, while also offering useful information for sports analysis and predictions.

Precise sports forecasting is a crucial competency for professional coaches, enabling them to formulate effective training methodologies and strategic competitive approaches. Conventional approaches sometimes employ intricate mathematical statistical techniques to enhance predictability; nevertheless, these methods are often constrained by dataset size and struggle with long-term forecasts involving variable distributions, particularly poorly in forecasting point-set-game multi-level matches. To address this difficulty, H. Liu et al. (2024) propose TM^2, a TCDformer-based Momentum Transfer Model for long-term sports prediction, which encompasses a momentum encoding module and a prediction module based on momentum transfer. Their analysis illustrates that TM^2 markedly surpasses current sports prediction models, achieving a 61.64% reduction in MSE and a 63.64% decrease in MAE, thus establishing a new standard in sports event prediction.

Another hybrid AI model idea was introduced by C. Liu et al. (2025), which integrated ARIMA and LSTM methodologies to conduct time series forecasting, specifically targeting the modeling of national medal counts to yield insights into prospective Olympic results. They achieved the final converged outcomes of RMSE and MAE values with consistent training cycles, specifically 0.098 for RMSE and 0.072 for MAE, which they assert are excellent results, demonstrating strong compatibility between ARIMA and LSTM in their hybrid AI model. Furthermore, their model may achieve an accuracy of up to 83% when utilizing a substantial volume of data. This demonstrated that their model is proficient and dependable for usage in this type of sports application.

The multidisciplinary characteristics of sports, along with numerous systemic and non-systemic components, complicate the prediction of match outcomes when relying on a singular disciplinary perspective. Consequently, Hsu (2021) conducted a study employing a DL methodology utilizing CNN and logistic regression for time series modelling to forecast sports match results. The primary data used in the prediction are sourced from betting market odds and real scores of each National Football League (NFL) game and are subsequently converted into sports candlesticks. The predictive efficacy of their model, CNN, was significantly enhanced by the proposed two-stage methodology, which utilized a logistic regression judgment model, achieving a prediction accuracy of 69% and surpassing that of the betting market.

Despite the valuable outcomes of data-driven algorithms in sports injury risk assessment, insufficient model generalization and the inability to automate feature extraction have hindered the practical application of study findings. Consequently, Ye et al. (2023) endeavor to develop an injury risk prediction model for running by integrating time-series picture encoding with DL techniques to tackle this issue more effectively. In comparison to the optimal model in their research, the proposed model shows enhancements in area under the curve (AUC), G_{mean}, sensitivity, and specificity by 23.9, 27.5, 39.7, and 16.2%, respectively, thus affirming the model's validity and applicability in injury risk prediction. The model presented in this paper provides a robust tool for predicting injury risk, aiming to enhance future sports injury prevention efforts.

A multitude of learning-based methodologies for efficient human activity recognition (HAR) have been recently established. Wearable inertial sensors are essential for HAR research to analyze sport-related activities. Sports activities are inherently unpredictable and have traditionally been categorized as complex, with standard ML methods employed to address HAR challenges. A hybrid DL model, designated as a multimodal bidirectional gated recurrent unit (MBiGRU) neural network, was introduced by Mekruksavanich and Jitpattanakul (2022) to identify commonplace sport-related actions, employing the publicly available Union Cycliste Internationale and Sports Dataset as a benchmark to evaluate the model's efficacy relative to other DL architectures in their research. The experimental findings suggested that the MBiGRU model achieved an accuracy rate of 99.55% and an F1-score of 99.55%, surpassing the baseline DL models in their study. Their proposed approach enhanced the efficacy of sensor-based HAR by utilizing the suggested MBiGRU paradigm.

Table 3 presents the best method among all those used in DL for predicting sports-related activities by implementing time series prediction, real-time prediction, or capturing temporal dependencies.

Summary

The entire journal above covered all research publications and journals that dealt with the application of DL techniques to the prediction of time series or real-time predictions for sports applications. The period covered by the paper examined spans from 2020 to 2025. Like in ML technique selection, a thorough examination of the application in sports, the performance of the AI models, and the outcomes that they achieved were the primary considerations that led to the selection of the findings from the comparison table and same as ML selection every one of the findings from the comparison table of DL will also serve as a guide for the subsequent study that will be conducted on this article, which is the construction of the Peephole-GRU Predictor for cyclist player performance prediction. A further observation that can be made is that between 2020 and 2025, a huge number

Table 3
Comparison of the best method by previous researchers based on the results obtained for implementing time series prediction and real-time prediction in deep learning methods

References	Algorithm method	Result			
		Model concept	Accuracy	Limitation	Application
Hilmkil et al. (2018)	LSTM	Vanila	Final min, mean, and max RMSE of 2.51, 5.62, and 25.67 on the validation set	Time-consuming, which prevents them from exploring all modifications to the learning parameters	Predict the heart-rate response of a cyclist during a training session
Zhang et al. (2022)	Attention-based sports competition performance prediction model (AS-LSTM)	Hybrid	The accuracy of the model reached 80% during its performance evaluation	A more appropriate sliding time window, a more effective word embedding method, and more accurate attention to athletes' psychological and physical states	Predict the sports competition results in a football match
Sadr et al. (2025)	LSTM	Vanila	Achieved the highest accuracy at 91.5%	The models require large and high-quality datasets for training, and incomplete or noisy data could significantly affect model performance	Predicting sports injuries for managing athletes' performance and health
Tang et al. (2025)	SAC-optimized DL model	Hybrid	Data processing accuracy (98.5%) in real-time monitoring	Heavily reliant on high-performance hardware, which may limit its broader applicability in diverse environments, and the system may still encounter some errors and delays, indicating the need for further algorithm optimization in future research	Focuses on real-time monitoring and analysis of track and field athletes

Note. LSTM = Long short-term memory; RMSE = Root mean square error; AS-LSTM = attention-based sports-aware Long short-term memory; SAC = Soft actor–critic; DL = Deep learning

of researchers conducted series prediction or real-time prediction using the DL method. Furthermore, a huge number of researchers were also focused on a hybrid model of AI, rather than using the vanilla type. This may be caused by parameters such as accuracy and F1 score, which produce much higher values and a higher success rate in predicting sports elements than using a vanilla type of algorithm. The study by Zhang et al. (2022) became one of the main references (apart from the ML method) for our research. This is because their method of using the AS-LSTM model, which they have created and modified from the vanilla LSTM method, is one of the creative and innovative approaches to predicting sports competition results in football matches, thereby helping sports teams train and improve. This method was chosen from the best of the five models in the comparison table. The reason for this is not only that the model they have adapted is highly suited and compatible with their parameters, but also that the data they provide, which demonstrates the model's accuracy, reached 80% during their performance evaluation, further enhancing our confidence in their empirical research. Because this paper employs a hybrid method in the LSTM algorithm, it is also relevant to our future research, which will involve modifying an LSTM model into a GRU and a Peephole model. Therefore, this paper is a suitable reference for our future research.

CONCLUSION

Sports analytics has been transformed by the inclusion of AI, especially ML and DL, into performance prediction, injury avoidance, and strategic decision-making. Particularly in cycling, basketball, and football, this paper presents a thorough study of time series and real-time prediction approaches used in various sports. Through an in-depth analysis of AI technologies from 2020 to 2025, both ML and DL models are crucial for enhancing sports analytics by providing more precise and data-driven insights. One important result of this study is the significant influence of AI on predictive modelling-based athlete performance optimization. High accuracy in forecasting outcomes and evaluating several performance criteria has been demonstrated by methods such as SVM, gradient boosting, LSTM networks, and hybrid models. Time series analysis enables athletes and coaches to make informed decisions based on real-time feedback, thereby guiding their choices of training and competition plans. AI applications, including VO_2 prediction, biomechanical analysis, and race strategy modelling, have given a competitive edge in cycling. The results show that hybrid AI architecture, LSTM-based models, and the LM + RM + HR approach all contribute to improved performance and reduced injury risk. Basketball and football have also seen similar developments, as AI-driven decision support systems enhance tactical planning, player scouting, and the prediction of game outcomes. There are still difficulties even with these developments. Using AI in sports requires large-scale databases, robust computing tools, and the ability to extend forecasts across multiple athletic

fields. Furthermore, the complexity of human performance introduces variability that AI models must constantly adjust to, emphasizing the necessity of further improvement and hybrid approaches that combine conventional sports science with AI-powered analytics. Furthermore, extending real-time AI applications from top-notch sports to grassroots levels may democratize performance statistics and provide amateur athletes and coaches with valuable insights. This demonstrates how quickly AI has become a transformative tool in sports analytics, as it provides unmatched performance prediction and clarity in strategic decision-making. The ongoing development of AI-driven techniques guarantees a data-driven approach to training, competition, and injury prevention, redefining how sports professionals evaluate and maximize performance. From all these studies, it is clear that every previous technique from past researchers will support and serve as a starting point for the development of the hybrid peephole-GRU predictor in this research, aiming to improve predictive capabilities in cycling sports.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support from the Fundamental Research Grant Scheme, under grant number FRGS/1/2024/ICT06/UNIMAP/02/1, from the Ministry of Higher Education Malaysia.

REFERENCES

Aarons, M. F., Young, C. M., Bruce, L., & Dwyer, D. B. (2023). Real time prediction of match outcomes in Australian football. *Journal of Sports Sciences*, *41*(11), 1115–1125. https://doi.org/10.1080/02640414.2023.2259266

Ahmadi, R., Rasoulian, S., Veisari, S. F., Parsaei, A., Heidary, H., Herzog, W., & Komeili, A. (2024). A machine learning approach for predicting pedaling force profile in cycling. *Sensors*, *24*(19), 6440. https://doi.org/10.3390/s24196440

Aleza, M. E., & Vetrithangam, D. (2023). Use of artificial intelligence to avoid errors in referring a football match. In *International Conference on Artificial Intelligence and Applications Alliance Technology Conference* (pp. 1–6). IEEE. https://doi.org/10.1109/ICAIA57370.2023.10169463

Alireza. (n.d.). *Radial basis function neural networks (with parameter selection using K-means)*. MathWorks. https://www.mathworks.com/matlabcentral/fileexchange/52580-radial-basis-function-neural-networks-with-parameter-selection-using-k-

Almaleck, P., Massucco, S., Mosaico, G., Saviozzi, M., Serra, P., & Silvestro, F. (2024). Electrical consumption forecasting in sports venues: A proposed approach based on neural networks and ARIMAX Models. *Sustainable Cities and Society*, *100*, 105019. https://doi.org/10.1016/j.scs.2023.105019

Almarashi, A. M., Daniyal, M., & Jamal, F. (2024). A novel comparative study of NNAR approach with linear stochastic time series models in predicting tennis player's performance. *BMC Sports Science, Medicine and Rehabilitation*, *16*, 28. https://doi.org/10.1186/s13102-024-00815-7

Amazon Web Services. (n.d.). *What is a neural network?* AWS. https://aws.amazon.com/what-is/neural-network/#:~:text=A%20neural%20network%20is%20a,that%20resembles%20the%20human%20brain.

Amendolara, A., Pfister, D., Settelmayer, M., Shah, M., Wu, V., Donnelly, S., Johnston, B., Peterson, R., Sant, D., Kriak, J., & Bills, K. (2023). An overview of machine learning applications in sports injury prediction. *Cureus: Journal of Medical Science*, *15*(9), e46170. https://doi.org/10.7759/cureus.46170

Anderson, J., & Rainie, L. (2023). *As AI spreads, experts predict the best and worst changes in digital life by 2035*. Pew Research Center. https://www.pewresearch.org/internet/2023/06/21/as-ai-spreads-experts-predict-the-best-and-worst-changes-in-digital-life-by-2035/?gad_source=1&gad_campaignid=22378837192&gbraid=0AAAAA-ddO9FyMKYfAtPXcL5wC6MOw4YCI&gclid=Cj0KCQjwrc7GBhCfARIsAHGcW5UC6B1lr3xK306xIWrAyon0_xnLrOiaQ8BCE8E7IQd6VokKSTzix28aAhtxEALw_wcB

Atkinson, G., Davison, R., Jeukendrup, A., & Passfield, L. (2003). Science and cycling: Current knowledge and future directions for research. *Journal of Sports Sciences*, *21*(9), 767–787. https://doi.org/10.1080/0264041031000102097

Bangsbo, J. (2014). Physiological demands of football. *Sports Science*, *27*(125), 1–6.

Bangsbo, J. (2015). Performance in sports – With specific emphasis on the effect of intensified training. *Scandinavian Journal of Medicine and Science in Sports*, *25*(S4), 88–99. https://doi.org/10.1111/sms.12605

Biró, A., Cuesta-Vargas, A. I., & Szilágyi, L. (2024). AI-assisted fatigue and stamina control for performance sports on IMU-generated multivariate times series datasets. *Sensors*, *24*(1), 132. https://doi.org/10.3390/s24010132

Britannica. (2025). *Bayesian analysis: Statistics*. https://www.britannica.com/science/Bayesian-analysis

Buchanan, R., Eliakim, R., & Eliakim, E. (2022). *Injury risk forecasting with Zone7 AI*. Zone7. https://zone7.ai/case-studies/validation-study/validation-study-injury-risk-forecasting-with-zone7-ai/

Calleja-González, J., Mallo, J., Cos, F., Sampaio, J., Jones, M. T., Marqués-Jiménez, D., Mielgo-Ayuso, J., Freitas, T. T., Alcaraz, P. E., Vilamitjana, J., Ibañez, S. J., Cuzzolin, F., Terrados, N., Bird, S. P., Zubillaga, A., Huyghe, T., Jukic, I., Lorenzo, A., Loturco, I., … Lago-Peñas, C. (2023). A commentary of factors related to player availability and its influence on performance in elite team sports. *Frontiers in Sports and Active Living*, *4*, 1077934. https://doi.org/10.3389/fspor.2022.1077934

Chen, X., & Yuan, G. (2021). Sports injury rehabilitation intervention algorithm based on visual analysis technology. *Mobile Information Systems*, *2021*, 9993677. https://doi.org/10.1155/2021/9993677

Cojocariu, I.-C. (2022). Predictive models applied in sports management – Literature review on research trends. *Journal of Public Administration, Finance and Law*, *11*(23), 148–153. https://doi.org/10.47743/jopafl-2022-23-11

Craig, N. P., & Norton, K. I. (2001). Characteristics of track cycling. *Sports Medicine*, *31*, 457–468. https://doi.org/10.2165/00007256-200131070-00001

Cui, J., Du, H., & Wu, X. (2023). Data analysis of physical recovery and injury prevention in sports teaching based on wearable devices. *Preventive Medicine*, *173*, 107589. https://doi.org/10.1016/j.ypmed.2023.107589

Dambroz, F., Clemente, F. M., & Teoldo, I. (2022). The effect of physical fatigue on the performance of soccer players: A systematic review. *PLOS One*, *17*(7), e0270099. https://doi.org/10.1371/journal.pone.0270099

De Beéck, T. O., Meert, W., Schütte, K., Vanwanseele, B., & Davis, J. (2018). Fatigue prediction in outdoor runners via machine learning and sensor fusion. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 606–615). Association for Computing Machinery. https://doi.org/10.1145/3219819.3219864

Di Salvo, V., Adam, C., Barry, M., & Marco, C. (2006). Validation of Prozone®: A new video-based performance analysis system. *International Journal of Performance Analysis in Sport*, *6*(1), 108–119. https://doi.org/10.1080/24748668.2006.11868359

Dolci, F., Hart, N. H., Kilding, A. E., Chivers, P., Piggott, B., & Spiteri, T. (2020). Physical and energetic demand of soccer: A brief review. *Strength and Conditioning Journal*, *42*(3), 70–77. https://doi.org/10.1519/SSC.0000000000000533

Eid, A. I. A., Miled, A. B., Fatnassi, A., Nawaz, M. A., Mahmoud, A. F. A., Abdalla, F. A., Jabnoun, C., Dhibi, A., Allan, F. M., Elhossiny, M. A., Belhaj, S., & Mohamed, I. B. (2024). Sports prediction model through cloud computing and big data based on artificial intelligence method. *Journal of Intelligent Learning Systems and Applications*, *16*(2), 53–79. https://doi.org/10.4236/jilsa.2024.162005

Elmagd, A. M. (2019). General psychological factors affecting physical performance and sports. *Journal of Advances in Sports and Physical Education*, *2*(7), 142–152. https://doi.org/10.36348/JASPE.2019.v02i07.004

Faria, E. W., Parker, D. L., & Faria, I. E. (2005a). The science of cycling: Physiology and training – Part 1. *Sports Medicine*, *35*, 285–312. https://doi.org/10.2165/00007256-200535040-00002

Faria, E. W., Parker, D. L., & Faria, I. E. (2005b). The science of cycling: Factors affecting performance - Part 2. *Sports Medicine*, *35*, 313–337. https://doi.org/10.2165/00007256-200535040-00003

Fédération Internationale de Football Association. (2022). *Semi-automated offside technology to be used at FIFA World Cup 2022™*. FIFA. https://inside.fifa.com/media-releases/semi-automated-offside-technology-to-be-used-at-fifa-world-cup-2022-tm

Fujii, K. (2021). Data-driven analysis for understanding team sports behaviors. *Journal of Robotics and Mechatronics*, *33*(3), 505–514. https://doi.org/10.20965/jrm.2021.p0505

Gavin, R. (2022). *Factors that effect sports performance: Sleep, mood, having a pre-match routine and romantic relationships* [Degree's thesis, National College of Ireland]. NORMA eResearch. https://norma.ncirl.ie/5641/1/rorygavin.pdf

GeeksforGeeks. (2025a). *Convolutional Neural Network (CNN) in machine learning*. https://www.geeksforgeeks.org/deep-learning/convolutional-neural-network-cnn-in-machine-learning/

GeeksforGeeks. (2025b). *Generative Adversarial Network (GAN)*. https://www.geeksforgeeks.org/deep-learning/generative-adversarial-network-gan/Retrieved (date) from url

GeeksforGeeks. (2025c). *Gradient boosting in ML*. https://www.geeksforgeeks.org/machine-learning/ml-gradient-boosting/

GeeksforGeeks. (2025d). *Support Vector Machine (SVM) algorithm*. https://www.geeksforgeeks.org/machine-learning/support-vector-machine-algorithm/

GeeksforGeeks. (2025e). *What is LSTM - Long Short Term Memory?* https://www.geeksforgeeks.org/deep-learning/deep-learning-introduction-to-long-short-term-memory/

Gomes, J., Portela, F., & Santos, M. F. (2013). Decision support system for predicting football game result. In X. Zhuang (Ed.), *Proceedings of the 19th International Conference on Computers, Recent Advance in Computer Science* (pp. 348-353). INASE Conferences. https://www.inase.org/library/2015/zakynthos/COMPUTERS.pdf

Gribble, A. (2021). *Browns introduce Express Access at FirstEnergy Stadium*. Cleveland Browns. https://www.clevelandbrowns.com/news/browns-introduce-express-access-at-firstenergy-stadium

Guo, H., Tang, R., Ye, Y., Li, Z., & He, X. (2017). *DeepFM: A factorization-machine based neural network for CTR prediction*. arXiv. https://doi.org/10.48550/arXiv.1703.04247

Hastings, G. (2023). *IBM brings generative AI commentary and AI draw analysis to the Wimbledon digital experience*. IBM. https://newsroom.ibm.com/2023-06-21-IBM-Brings-Generative-AI-Commentary-and-AI-Draw-Analysis-to-the-Wimbledon-Digital-Experience

He, X., & Chua, T.-S. (2017). Neural factorization machines for sparse predictive analytics. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 355–364). Association for Computing Machinery. https://doi.org/10.1145/3077136.3080777

Hilmkil, A., Ivarsson, O., Johansson, M., Kuylenstierna, D., & van Erp, T. (2018). *Towards machine learning on data from professional cyclists*. arXiv. https://doi.org/10.48550/arXiv.1808.00198

Hribernik, M., & Kos, A. (2023). Exploring the applicability of haptic actuators in aquatic environments. *Internet of Things*, *24*, 100924. https://doi.org/10.1016/j.iot.2023.100924

Hsu, Y.-C. (2021). Using convolutional neural network and candlestick representation to predict sports match outcomes. *Applied Sciences*, *11*(14), 6594. https://doi.org/10.3390/app11146594

Hu, X. (2024). Prediction of sports achievements based on time series analysis. In *International Conference on Cloud Computing, Performance Computing, and Deep Learning* (Vol. 13281, pp. 73–78). SPIE. https://doi.org/10.1117/12.3050728

Ibrahim, M. R., Haworth, J., Christie, N., Cheng, T., & Hailes, S. (2017). Cycling near misses: A review of the current methods, challenges and the potential of an AI-embedded system. *Transport Reviews*, *41*(3), 304–328. https://doi.org/10.1080/01441647.2020.1840456

Imran, H. A. (2022). Khail-Net: A shallow Convolutional Neural Network for recognizing sports activities using wearable inertial sensors. *IEEE Sensors Letters*, *6*(9), 1–4. https://doi.org/10.1109/LSENS.2022.3197396

International Business Machines Corporation. (2021). *What is linear regression?* IBM. https://www.ibm.com/think/topics/linear-regression

International Business Machines Corporation. (n.d.a). *What is random forest?* IBM. https://www.ibm.com/think/topics/random-forest#:~:text=Random%20forest%20is%20a%20commonly,both%20classification%20and%20regression%20problems.

International Business Machines Corporation. (n.d.b). *What is the k-nearest neighbors (KNN) algorithm?* IBM. https://www.ibm.com/think/topics/knn#:~:text=The%20k%2Dnearest%20nneighbor%20(KNN)%20algorithm%20is%20a%20non,used%20in%20macmach%20learning%20today.

Karetnikov, A. D. (2019). *Application of data-driven analytics on sport data from a professional bicycle racing team* [Master's thesis, Eindhoven University of Technology]. TU/e Research Portal. https://pure.tue.nl/ws/portalfiles/portal/139933495/ThesisAKaretnikov.pdf

Kolambe, M., & Arora, S. (2024). Forecasting the future: A comprehensive review of time series prediction techniques. *Journal of Electrical Systems*, *20*(2s), 575–586. https://doi.org/10.52783/jes.1478

Lastella, M., Roach, G. D., Halson, S. L., & Sargent, C. (2015). Sleep/wake behaviours of elite athletes from individual and team sports. *European Journal of Sport Science*, *15*(2), 94–100. https://doi.org/10.1080/17461391.2014.932016

Li, G. (2022). Construction of sports training performance prediction model based on a generative adversarial deep neural network algorithm. *Computational Intelligence and Neuroscience*, *2022*, 1211238. https://doi.org/10.1155/2022/1211238

Li, H., Cui, C., & Jiang, S. (2024). Strategy for improving the football teaching quality by AI and metaverse-empowered in mobile internet environment. *Wireless Networks*, *30*, 4343–4352. https://doi.org/10.1007/s11276-022-03000-1

Li, L., & Shi, X. (2014). Tennis "Hawk-Eye" technical research. *Journal of Chemical and Pharmaceutical Research*, *6*(6), 298–305.

Li, N., Hu, W., Ma, Y., & Xiang, H. (2024). Machine learning prediction of pulmonary oxygen uptake from muscle oxygen in cycling. *Journal of Sports Sciences*, *42*(14), 1299–1307. https://doi.org/10.1080/02640414.2024.2388996

Lingeswaran, S. (2023). *WSC Sports: How AI has led the highlights revolution*. Sportcal. https://www.sportcal.com/features/wsc-sports-how-ai-has-led-the-highlights-revolution/

Liu, C., Ma, C., & Zhou, X. (2025). *Exploring patterns behind sports*. arXiv. https://doi.org/10.48550/arXiv.2502.07491

Liu, H., Huang, X., Gu, J., Shi, J., He, N., & Feng, T. (2024). TCDformer-based momentum transfer model for long-term sports prediction. *Expert Systems with Applications*, *289*, 128310. https://doi.org/10.1016/j.eswa.2025.128310

Luo, R., & Krishnamurthy, V. (2023). *Who you play affects how you play: Predicting sports performance using graph attention networks with temporal convolution*. arXiv. https://doi.org/10.48550/arXiv.2303.16741

Luo, S., Soh, K. G., Zhao, Y., Soh, K. L., Sun, H., Nasiruddin, N. J. M., Zhai, X., & Ma, L. (2023). Effect of core training on athletic and skill performance of basketball players: A systematic review. *PLOS One*, *18*(6), e0287379. https://doi.org/10.1371/journal.pone.0287379

Mekruksavanich, S., & Jitpattanakul, A. (2022). Sport-related activity recognition from wearable sensors using bidirectional GRU Network. *Intelligent Automation and Soft Computing*, *34*(3), 1907–1925. https://doi.org/10.32604/iasc.2022.027233

Miah, J., Mamun, M., Rahman, M. M., Mahmud, M. I., Ahmad, S., & Nasir, M. H. (2022). MHfit: Mobile health data for predicting athletics fitness using machine learning models. In *2nd International Seminar on Machine Learning, Optimization, and Data Science* (pp. 584–589). IEEE. https://doi.org/10.1109/ISMODE56940.2022.10180967

Mullin, B. (2016). *The Associated Press will use automated writing to cover the minor leagues*. Poynter. https://www.poynter.org/tech-tools/2016/the-associated-press-will-use-automated-writing-to-cover-the-minor-leagues/

Nakisa, N., & Rahbardar, M. G., (2021). A review of the effect of psychological factors on soccer player's performance. *Indian Journal of Physical Education, Sports and Applied Science*, *11*(1), 87–102.

Noble, J. (2024). *What are ARIMA models?* IBM. https://www.ibm.com/think/topics/arima-model

Nokihara, Y., Hori, R., Hachiuma, R., & Saito, H. (2023). Prediction of shuttle trajectory in badminton using player's position. In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* (Vol. 5, 788–795). SciTePress. https://doi.org/10.5220/0011785800003417

NVIDIA Corporation. (n.d.). *XGBoost*. https://www.nvidia.com/en-us/glossary/xgboost/

Oytun, M., Tinazci, C., Sekeroglu, B., Acikada, C., & Yavuz, H. U. (2020). Performance prediction and evaluation in female handball players using machine learning models. *IEEE Access*, *8*, 116321–116335. https://doi.org/10.1109/ACCESS.2020.3004182

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*, 89. https://doi.org/10.1186/s13643-021-01626-4

Pantzalis, V. C., & Tjortjis, C. (2020). Sports analytics for football league table and player performance prediction. In *11th International Conference on Information, Intelligence, Systems and Applications* (pp. 1-8). IEEE. https://doi.org/10.1109/IISA50023.2020.9284352

Pappalardo, L., Luca, G., Alessio, R., & Paolo, C. (2020). *Explainable injury forecasting in soccer via multivariate time series and Convolutional Neural Networks*. Barca Innovation Hub. https://static.capabiliaserver.com/frontend/clients/barca/wp_prod/wp-content/uploads/2020/01/c6658839-paper-format-luca-pappalardo-1.pdf

Pennsylvania State University. (n.d.). *Introduction to GLMs*. https://online.stat.psu.edu/stat504/lesson/6/6.1

Phillips, K. E., & Hopkins, W. G. (2020). Determinants of cycling performance: A review of the dimensions and features regulating performance in elite cycling competitions. *Sports Medicine - Open*, *6*, 23. https://doi.org/10.1186/s40798-020-00252-z

Plakias, S., & Michailidis, Y. (2024). Factors affecting the running performance of soccer teams in the Turkish Super League. *Sports*, *12*(7), 196. https://doi.org/10.3390/sports12070196

Proulx, D., & Smith, D. J. (2014). *Track skills development grid*. https://cyclingbc.net/wp-content/uploads/2022/01/Cycling-Disciplines-Skills-Development-Grid.pdf

Ren, L., Wang, Y., & Li, K. (2024). Real-time sports injury monitoring system based on the deep learning algorithm. *BMC Medical Imaging*, *24*, 122. https://doi.org/10.1186/s12880-024-01304-6

Richman, P. (2018). *NBA 2013-14 Season: New SportVU system in arenas to improve player tracking*. Bleacher Report. https://bleacherreport.com/articles/1764278-nba-to-install-sportvu-system-in-arenas-for-2013-14-and-improve-player-tracking

Ruddock, J. (2024). *Spiideo unveils industry-first product updates underscoring significant advancement in AI-automated sports broadcasting*. SVG Europe. https://www.svgeurope.org/blog/news-roundup/spiideo-unveils-industry-first-product-updates-underscoring-significant-advancement-in-ai-automated-sports-broadcasting/

Sadr, M. M., Khani, M., & Tootkaleh, S. M. (2025). Predicting athletic injuries with deep learning: Evaluating CNNs and RNNs for enhanced performance and safety. *Biomedical Signal Processing and Control*, *105*, 107692. https://doi.org/10.1016/j.bspc.2025.107692

Salvi, M., Acharya, U. R., Molinari, F., & Meiburger, K. M. (2021). The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis. *Computers in Biology and Medicine*, *128*, 104129. https://doi.org/10.1016/j.compbiomed.2020.104129

Sarker, I. H. (2021). Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, *2*, 420. https://doi.org/10.1007/s42979-021-00815-1

Seedhouse, E. (2017). Spaceports around the world. In *Spaceports around the world, a global growth industry* (pp. 101–113). Springer. https://doi.org/10.1007/978-3-319-46846-4_9

Shukla, R., Jindal, P., Gupta, A., & Patil, H. Y. (2022). Total revenue prediction of a sports management application: Grook using machine learning models. In *13th International Conference on Computing Communication and Networking Technologies* (pp. 1–6). IEEE. https://doi.org/10.1109/ICCCNT54827.2022.9984472

Sim, J. (2024). *Global sports industry revenues to reach US$260bn by 2033*. SportsPro. https://www.sportspro.com/news/global-sports-industry-revenue-projection-2033-two-circles/#:~:text=Sports%20industry%20currently%20generates%20US$159%20billion%20in%20annual%20revenue,projected%20to%20outperform%20the%20market

Sports Business Journal. (2023). *Timeline of AI use in sports*. SBJ. https://www.sportsbusinessjournal.com/Articles/2023/12/04/ai-timeline/

Stetter, B. J., Ringhof, S., Krafft, F. C., Sell, S., & Stein, T. (2020). Estimation of knee joint forces in sport movements using wearable sensors and machine learning. *Sensors*, *19*(17), 3690. https://doi.org/10.3390/s19173690

Stübinger, J., Mangold, B., & Knoll, J. (2020). Machine learning in football betting: Prediction of match results based on player characteristics. *Applied Sciences*, *10*(1), 46. https://doi.org/10.3390/app10010046

Suo, X., Tang, W., & Li, Z. (2024). Motion capture technology in sports scenarios: A survey. *Sensors*, *24*(9), 2947. https://doi.org/10.3390/s24092947

Taboga, M. (2021). *Bayesian estimation of the parameters of the normal distribution*. StatLect. https://www.statlect.com/fundamentals-of-statistics/normal-distribution-Bayesian-estimation.

Tang, X., Long, B., & Zhou, L. (2025). Real-time monitoring and analysis of track and field athletes based on edge computing and deep reinforcement learning algorithm. *Alexandria Engineering Journal*, *114*, 136–146. https://doi.org/10.1016/j.aej.2024.11.024

Tingting, L. (2015). Research on the intelligent teaching system of college basketball based on artificial intelligence. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (18B), 49–60.

Vec, V., Tomažič, S., Kos, A., & Umek, A. (2024). Trends in real-time artificial intelligence methods in sports: A systematic review. *Journal of Big Data*, *11*, 148. https://doi.org/10.1186/s40537-024-01026-0

Wang, F., & Zheng, G. (2022). What are the changes in basketball shooting pattern and accuracy in National Basketball Association in the past decade? *Frontiers in Psychology*, *13*, 917980. https://doi.org/10.3389/fpsyg.2022.917980

Wang, G., & Ren, T. (2024). Design of sports achievement prediction system based on U-net convolutional neural network in the context of machine learning. *Heliyon*, *10*(10), e30055. https://doi.org/10.1016/j.heliyon.2024.e30055

Wang, G., Pu, P., & Shen, T. (2020). An efficient gene bigdata analysis using machine learning algorithms. *Multimedia Tools and Applications*, *79*, 9847–9870. https://doi.org/10.1007/s11042-019-08358-7

Wang, L., Zhang, H., & Yuan, G. (2021). Big data and deep learning-based video classification model for sports. *Wireless Communications and Mobile Computing*, *2021*, 1140611. https://doi.org/10.1155/2021/1140611

Watson, N., Hendricks, S., Stewart, T., & Durbach, I. (2021). Integrating machine learning and decision support in tactical decision-making in rugby union. *Journal of the Operational Research Society*, *72*(10), 2274–2285. https://doi.org/10.1080/01605682.2020.1779624

WHOOP. (2023). *WHOOP unveils the new WHOOP coach powered by OpenAI*. https://www.whoop.com/us/en/thelocker/whoop-unveils-the-new-whoop-coach-powered-by-openai/?srsltid=AfmBOorahCjNdPWXNgAA4VFS50SW73Z1XfwFg1Crzfc3tE37o43LzL22

Wohlwend, B. (2023). *Three regression models for data science: Linear regression, lasso regression, and ridge regression*. Medium. https://medium.com/@brandon93.w/three-regression-models-for-data-science-linear-regression-lasso-regression-and-ridge-regression-6aac73c0d7a5

Wood, R. (2020). *The many sports of cycling*. Topend Sports. https://www.topendsports.com/sport/cycling-sports.htm

Wright, C., Atkins, S., Jones, B., & Todd, J. (2013). The role of performance analysts within the coaching process: Performance analysts survey 'The role of performance analysts in elite football club settings.'. *International Journal of Performance Analysis in Sport*, *13*(1), 240–261. https://doi.org/10.1080/24748668.2013.11868645

Wu, Y., Ma, Z., Zhao, H., Li, Y., & Sun, Y. (2020). Achieve personalized exercise intensity through an intelligent system and cycling equipment: A machine learning approach. *Applied Sciences*, *10*(21), 7688. https://doi.org/10.3390/app10217688

Xu, J. (2022). Prediction and planning of sports competition based on deep neural network. *Computational Intelligence and Neuroscience*, *2022*, 1906580. https://doi.org/10.1155/2022/1906580

Xu, T., & Tang, L. (2021). Adoption of machine learning algorithm-based intelligent basketball training robot in athlete injury prevention. *Frontiers in Neurorobotics*, *14*, 620378. https://doi.org/10.3389/fnbot.2020.620378

Yan, W., Jiang, X., & Liu, P. (2023). A review of basketball shooting analysis based on artificial intelligence. *IEEE Access*, *11*, 87344–87365. https://doi.org/10.1109/ACCESS.2023.3304631

Ye, X., Huang, Y., Bai, Z., & Wang, Y. (2023). A novel approach for sports injury risk prediction: Based on time-series image encoding and deep learning. *Frontiers in Physiology*, *14*, 1174525. https://doi.org/10.3389/fphys.2023.1174525

Yuan, X. (2024). Higher dimensional sports statistics and real-time game prediction. *Advances in Engineering Innovation*, *8*, 9–18. https://doi.org/10.54254/2977-3903/8/2024071

Zhang, Q., Zhang, X., Hu, H., Li, C., Lin, Y., & Ma, R. (2022). Sports match prediction model for training and exercise using attention-based LSTM network. *Digital Communications and Networks*, *8*(4), 508–515. https://doi.org/10.1016/j.dcan.2021.08.008

Zhao, J., & Li, G. (2023). A combined deep neural network and semi-supervised clustering method for sports injury risk prediction. *Alexandria Engineering Journal*, *80*, 191–201. https://doi.org/10.1016/j.aej.2023.08.048

ZhaoriGetu, H. (2022). Prediction of sports performance combined with deep learning model and analysis of influencing factors. *Scientific Programming*, *2022*, 4082906. https://doi.org/10.1155/2022/4082906

Zheng, D., & Yuan, Y. (2022). Time series data prediction and feature analysis of sports dance movements based on machine learning. *Computational Intelligence and Neuroscience*, *2022*, 5611829. https://doi.org/10.1155/2022/5611829

Zhou, Q. (2022). Sports achievement prediction and influencing factors analysis combined with deep learning model. *Scientific Programming*, *2022*, 3547703. https://doi.org/10.1155/2022/3547703

Zhu, Z. (2024). High dimensional sports statistics and machine learning in NBA. *Advances in Engineering Innovation*, *11*, 78–94. https://doi.org/10.54254/2977-3903/11/2024117

Pertanika J. Sci. & Technol. 33 (6): 2723 - 2746 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Comparative Study of Nonlinear Regression Models for Cleanliness Factor Prediction in Coal-Fired Utilities

Nur Aina Mohammad Abdul Aziz[1], Azura Che Soh[1]*, Asnor Juraiza Ishak[1], Nor Mohd Haziq Norsahperi[1] and Amir Redzuan Mohd Ibrahim[2]

[1]Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

[2]Sultan Azlan Shah Power Station, Tenaga Nasional Berhad, Teluk Rubiah, 32040 Seri Manjung, Perak, Malaysia

ABSTRACT

This study presents a comparative analysis of nonlinear regression models integrated with feature selection for predicting the cleanliness factor (CF) in coal-fired utilities. The models evaluated are regression trees (RT), support vector regression (SVR), ensembles of trees, and artificial neural networks (ANN). Different boiler designs introduce various operational parameters that influence cleanliness, making it more challenging to predict real-time data accurately. To enhance the model's predictive accuracy, the minimum redundancy maximum relevance (MRMR) feature selection technique was integrated, facilitating improved model performance by selecting the best subsets of variables. Model performance was assessed accordingly, where the number of selected features varies between 138 and 10. The results indicate that a combination of bagged trees and MRMR with 10 features achieved R^2 values of 0.973 for the training set and 0.976 for the test set, with a mean squared error (MSE) of 0.001 for both datasets. Compared to SVR and ANN, bagged trees consistently demonstrated superior predictive accuracy with reduced computational complexity. These findings confirm that ensemble-based models, particularly bagged trees with MRMR, provide the most accurate and computationally efficient approach for CF prediction. An accurate CF prediction creates more reliable information for a data-driven approach that solves the soot-blowing operational system. The system has the risk of either underblowing or overblowing steam during boiler cleaning. This risk, if not properly handled, may lead to more severe ash fouling and slagging issues, such as emergency shutdowns, metal corrosion, and declining heat transfer efficiency in coal-fired utilities. Overall, improving real-time boiler monitoring minimizes steam waste during soot-blowing operations.

Keywords: Cleanliness factor, coal-fired boiler, feature selection, machine learning, nonlinear regression, soot blowing optimization

ARTICLE INFO

Article history:
Received: 11 March 2025
Accepted: 04 July 2025
Published: 14 October 2025

DOI: https://doi.org/10.47836/pjst.33.6.14

E-mail addresses:
gs69737@student.upm.edu.my (Nur Aina Mohammad Abdul Aziz)
azuracs@upm.edu.my (Azura Che Soh)
asnorji@upm.edu.my (Asnor Juraiza Ishak)
nmhaziq@upm.edu.my (Nor Mohd Haziq Norsahperi)
amir.redzuan@tnb.com.my (Amir Redzuan Mohd Ibrahim)
* Corresponding author

© Universiti Putra Malaysia Press

INTRODUCTION

Coal-fired power plants generate substantial soot, leading to issues such as ash fouling and slagging in boiler sections (Wei et al., 2020). These issues not only reduce heat transfer efficiency but can also cause damage to boiler components, raising operational costs and creating additional maintenance demands. Soot-blowing mechanisms are employed to mitigate these effects, using high-temperature, high-pressure steam to remove soot from boiler walls and pipes (Kumari & Srinivasan, 2019). However, soot-blowing traditionally follows a fixed schedule, relying on operator experience, which can result in inconsistent cleaning due to human error (P. Li et al., 2023; Q. Li et al., 2020; Shi et al., 2021; Wen et al., 2022). To evaluate cleaning effectiveness, the CF is used as a metric, comparing the boiler's current condition to an ideally clean state (Shi et al., 2022).

A coal-fired power plant is known to operate with a high-dimensional dataset (Menn & Chudnovsky, 2021). Despite improvements in boiler design and technology, research on identifying key parameters that affect boiler cleanliness has been limited, particularly in utilizing data mining techniques for predictive modeling. Thota and Syed (2024) mentioned that while coal-fired utilities have much operational data, the usage of unwanted characteristics and past data is not being addressed properly, which later causes the prediction performance to degrade. Different boiler designs and configurations will introduce varying parameters that can impact the cleanliness of the boiler. Thus, relying on the expert advice based on previous boilers may overlook hidden correlations within the boiler.

This study addresses the need for feature selection methods to identify the most relevant parameters for accurate CF predictions. Feature selection algorithms help identify the best subsets of variables that enhance model performance by focusing on critical factors and reducing complexity (Bezerra et al., 2024; Jemai & Zarrad, 2023). Given that the CF ranges continuously from 0 to 1, regression learners are well-suited for this task, as they can capture continuous outputs based on the model's input features. Furthermore, considering the high-dimensional dataset, nonlinear relationships will be utilized to identify patterns among key parameters that influence boiler cleanliness.

Therefore, this research explores various nonlinear regression models that incorporate feature selection techniques, comparing their performance to determine the most effective approach for analyzing data from coal-fired boilers. The objective is to develop robust regression models that use real production data to predict the CF within the convective sections of coal-fired power plants, ultimately enhancing cleaning efficiency and operational stability.

Overview of Coal-Fired Boiler

The overview of an ultra-supercritical boiler power plant operated by one of the major utility providers in Malaysia is presented in Figure 1. The plant includes several key components: the boiler system, furnace system, boiler clean-up and start-up system, and

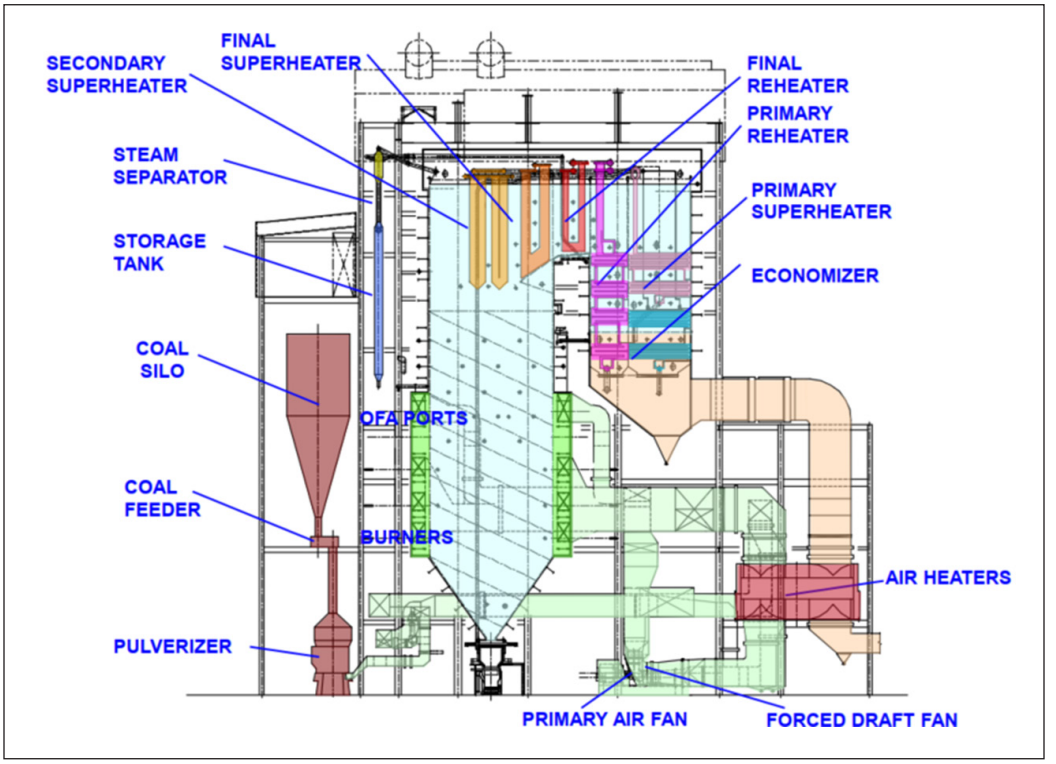


Figure 1. Overview of ultra-supercritical boiler power plant (*Source:* Tenaga Nasional Berhad [TNB], 2024)

air and gas system. Additionally, the metal temperature throughout the power plant is considered. The components are primarily categorized based on their respective heat exchangers.

The boiler system's pulverizer consists of six coal flow storage units, as listed in Table 1. Of these, five will be operated simultaneously, while one will serve as a backup for the coal flow. Pulverizers grind coal into a fine powder, which is subsequently used as fuel for combustion in the boiler to optimize combustion efficiency.

Table 1
Parameters of pulverizer

No.	Parameter	Unit
1.	Pulv A coal flow	t/h
2.	Pulv B coal flow	t/h
3.	Pulv C coal flow	t/h
4.	Pulv D coal flow	t/h
5.	Pulv E coal flow	t/h
6.	Pulv F coal flow	t/h

Note. Pulv = Pulverizer

The following parameters are essential for the main soot-blowing operation system, as shown in Table 2. The main steam pressure, flow, and temperature sensors measure the pressure, flow, and temperature generated within the boiler. An optical pyrometer sensor is used to monitor the flue gas components. Two types of fans are utilized: the primary air fan (PAF), which transports pulverized coal from the pulverizers to the furnace area, and the

forced draft fan (FDF), which supplies fresh air to the furnace for the combustion process. Additionally, feedwater pressure and flow sensors monitor the feedwater as it enters the boiler. The target load measures the power output required from the steam turbine, while the steam blow-off ring (SBR) steam flow sensor measures steam flow as part of the power plant's safety and operational systems.

The economizer is a component of the convective section of the boiler used to preheat feedwater before it enters the boiler. This process reduces the amount of coal required to produce high heat for steam generation. Additionally, the economizer cools down the flue gas before it is released into the atmosphere. Sensors monitor parameters, including the oxygen percentage, gas inlet and outlet temperatures, and feedwater inlet and outlet temperatures within the economizer. The detailed parameters for the economizer are listed in Table 3.

Table 2
Parameters of the main operation

No.	Parameter	Unit
1.	Main steam pressure	MPag
2.	Main steam flow	t/hr
3.	Main steam temperature 1	degC
4.	Optical pyrometer sensor 1	degC
5.	Optical pyrometer sensor 2	degC
6.	PAF A air volume flow	t/h
7.	PAF B air volume flow	t/h
8.	FDF A air volume flow	t/h
9.	FDF B air volume flow	t/h
10.	Feed water pressure	MPa
11.	FW flow	t/h
12.	Main steam (A) pressure	Mpag
13.	Main steam (B) pressure	MPag
14.	Target load (net)	MW
15.	SBR steam flow	t/h

Note. PAF = Primary air fan; FDF = Force draft fan; FW = Feedwater; SBR = Steam blow-off ring

Table 3
Parameters of economizer

No.	Parameter	Unit
1.	ECON A out gas O_2 anal	%
2.	ECON B out gas O_2 anal	%
3.	ECON out gas TEMP 1	degC
4.	ECON out gas TEMP 2	degC
5.	ECON inlet FW TEMP	degC
6.	ECON outlet FW TEMP	degC
7.	ECON B outlet FW TEMP	degC
8.	ECON inlet FW flow 3	t/h

Note. ECON = Economizer; O_2 = Oxygen; TEMP = Temperature; FW = Feedwater

The primary superheater (PSH), located within the convective section of the boiler, serves as the initial stage in the superheating process. Its primary function is to elevate the temperature of the steam without increasing its pressure, optimizing thermal efficiency while maintaining safe operational limits. Key parameters monitored by sensors include steam inlet and outlet temperatures, as well as spray water flow, ensuring effective and controlled temperature increases. The detailed parameters for the PSH are listed in Table 4.

The secondary superheater (SSH), also situated in the convective section, is the second stage of superheating and further elevates the steam's temperature to ultra-high levels. This additional temperature increase enhances the thermal energy available for power generation. Similar to the PSH, sensors monitor critical parameters such as the steam inlet and outlet

temperatures and spray water flow, ensuring precise temperature control at this advanced heating stage. The detailed parameters for the SSH are listed in Table 5.

Following the SSH, the final superheater (FSH), the third stage of superheating, maximizes the steam's temperature just before it enters the high-pressure turbine. This maximization step is crucial for achieving the highest possible efficiency in power generation. Sensors in the FSH monitor vital parameters, including steam inlets and outlet temperatures, to ensure optimal performance and safety. The detailed parameters for the FSH are listed in Table 6.

The reheater (RH) component, part of the boiler's convective section, reintroduces heat to the steam after it exits the high-pressure turbine, raising its temperature before it enters the low and intermediate-pressure turbines. This reheating process is essential for sustaining efficiency and preventing moisture formation, which could damage turbine blades. Key parameters, such as steam inlet and outlet temperatures, are continuously monitored by sensors to regulate this reheat phase effectively. The detailed parameters for the RH are listed in Table 7.

Table 4
Parameters of primary superheaters (PSH)

No.	Parameter	Unit
1.	PSH A inlet steam TEMP	degC
2.	PSH B inlet steam TEMP	degC
3.	PSH A out steam TEMP	degC
4.	PSH B out steam TEMP	degC
5.	PSH spray water flow (1st stage)	t/h

Note. TEMP = Temperature

Table 5
Parameters of secondary superheaters (SSH)

No.	Parameter	Unit
1.	SSH A inlet steam TEMP	degC
2.	SSH B inlet steam TEMP	degC
3.	SSH A out steam TEMP	degC
4.	SSH B out steam TEMP	degC
5.	SSH spray water CV A POS	%
6.	SSH spray water CV B POS	%
7.	SSH spray water flow	t/h

Note. TEMP = Temperature; CV = Control valve; POS = Position

Table 6
Parameters of final superheaters (FSH)

No.	Parameter	Unit
1.	FSH A inlet steam TEMP	degC
2.	FSH B inlet steam TEMP	degC
3.	FSH A out steam TEMP	degC
4.	FSH B out steam TEMP	degC

Note. TEMP = Temperature

Table 7
Parameters of reheaters (RH)

No.	Parameter	Unit
1.	PRH steam TEMP A	degC
2.	PRH steam TEMP B	degC
3.	FRH A out steam TEMP	degC
4.	FRH B out steam TEMP	degC
5.	HRH A steam pressure (A)	MPa
6.	HRH B steam pressure (B)	MPa
7.	CRH steam TEMP A	degC
8.	CRH steam TEMP B	degC
9.	CRH steam A pressure	MPa
11.	RH spray water CV A POS	%
12.	RH spray water CV B POS	%
13.	RH spray water flow	t/h
14.	HORIZ RH out gas TEMP 1	degC

Note. PRH = Primary reheater; FRH = Final reheater; HRH = Hot reheat; CRH = Cold reheat; HORIZ = Horizontal; TEMP = Temperature; CV = Control valve; POS = Position

The air heater (AH), located in the convective section of the boiler, preheats the combustion air by utilizing residual heat from the flue gases, thereby enhancing overall combustion efficiency. By raising the temperature of the incoming air, the AH reduces the energy required for combustion, thereby improving the boiler's thermal efficiency and reducing fuel consumption. Sensors within the AH monitor key parameters, such as the inlet gas temperature, to ensure optimal heat transfer and maintain efficient preheating conditions. This monitoring enables precise adjustments, maximizing both energy conservation and combustion efficiency. The detailed parameters for the AH are listed in Table 8.

Metal temperature sensors make up the largest number of sensors in the power plant, serving to monitor the temperature of various metal components. Due to the extensive area required for heat transfer, numerous sensors are installed within the boiler. Accurate temperature control is imperative for optimizing coal power plant efficiency. The metal temperature sensors are listed in Table 9.

Table 8
Parameters of air heaters (AH)

No.	Parameter	Unit
1.	AH A inlet gas TEMP	degC
2.	AH B inlet gas TEMP	degC

Note. TEMP = Temperature

Table 9
Metal temperature sensors

No.	Furnace	Superheaters	Reheaters
1.	MNJ:05:_BSC1_AI_0186	MNJ:05:_BSC2_AI_0281	MNJ:05:_BSC2_AI_0317
2.	MNJ:05:_BSC1_AI_0187	MNJ:05:_BSC2_AI_0282	MNJ:05:_BSC2_AI_0319
3.	MNJ:05:_BSC1_AI_0188	MNJ:05:_BSC2_AI_0283	MNJ:05:_BSC2_AI_0322
4.	MNJ:05:_BSC1_AI_0189	MNJ:05:_BSC2_AI_0284	MNJ:05:_BSC2_AI_0324
5.	MNJ:05:_BSC1_AI_0190	MNJ:05:_BSC2_AI_0285	MNJ:05:_BSC2_AI_0318
6.	MNJ:05:_BSC1_AI_0191	MNJ:05:_BSC2_AI_0286	MNJ:05:_BSC2_AI_0321
7.	MNJ:05:_BSC1_AI_0193	MNJ:05:_BSC2_AI_0287	MNJ:05:_BSC2_AI_0323
8.	MNJ:05:_BSC1_AI_0194	MNJ:05:_BSC2_AI_0289	MNJ:05:_BSC2_AI_0325
9.	MNJ:05:_BSC1_AI_0195	MNJ:05:_BSC1_AI_0165	MNJ:05:_BSC2_AI_0326
10.	MNJ:05:_BSC1_AI_0234	MNJ:05:_BSC1_AI_0166	MNJ:05:_BSC2_AI_0327
11.	MNJ:05:_BSC1_AI_0235	MNJ:05:_BSC1_AI_0167	MNJ:05:_BSC2_AI_0329
12.	MNJ:05:_BSC1_AI_0236	MNJ:05:_BSC1_AI_0244	MNJ:05:_BSC2_AI_0330
13.	MNJ:05:_BSC1_AI_0237	MNJ:05:_BSC1_AI_0245	MNJ:05:_BSC2_AI_0331
14.	MNJ:05:_BSC1_AI_0238	MNJ:05:_BSC1_AI_0246	MNJ:05:_BSC2_AI_0332
15.	MNJ:05:_BSC1_AI_0239	MNJ:05:_BSC1_AI_0259	MNJ:05:_BSC2_AI_0333
16.	MNJ:05:_BSC1_AI_0241	MNJ:05:_BSC1_AI_0260	MNJ:05:_BSC2_AI_0334
17.	MNJ:05:_BSC1_AI_0242	MNJ:05:_BSC1_AI_0261	
18.	MNJ:05:_BSC1_AI_0243	MNJ:05:_BSC1_AI_0262	
19.	MNJ:05:_BSC1_AI_0169	MNJ:05:_BSC1_AI_0263	
20.	MNJ:05:_BSC1_AI_0170	MNJ:05:_BSC1_AI_0265	

Table 9 *(continue)*

No.	Furnace	Superheaters	Reheaters
21.	MNJ:05:_BSC1_AI_0171	MNJ:05:_BSC1_AI_0266	
22.	MNJ:05:_BSC1_AI_0172	MNJ:05:_BSC1_AI_0310	
23.	MNJ:05:_BSC1_AI_0173	MNJ:05:_BSC2_AI_0308	
24.	MNJ:05:_BSC1_AI_0217	MNJ:05:_BSC2_AI_0309	
25.	MNJ:05:_BSC1_AI_0218	MNJ:05:_BSC2_AI_0310	
26.	MNJ:05:_BSC1_AI_0219	MNJ:05:_BSC2_AI_0311	
27.	MNJ:05:_BSC1_AI_0220	MNJ:05:_BSC2_AI_0313	
28.	MNJ:05:_BSC1_AI_0221	MNJ:05:_BSC2_AI_0314	
29.		MNJ:05:_BSC2_AI_0315	
30.		MNJ:05:_BSC2_AI_0316	

Note. MNJ:05: BSC1_AI_xxxx = Manjung unit 5 boiler superheater coil 1 analog input no. xxxx; MNJ:05: BSC2_AI_xxxx = Manjung unit 5 boiler superheater coil 2 analog input no. xxxx

METHODOLOGY

This research focuses on developing nonlinear regression models, enhanced with feature selection techniques, to analyze and identify the most robust model for gaining insights into new parameters within the convective sections of the boiler. The model development process begins with data collection from the power plant, followed by data preprocessing to ensure the quality and accuracy of the data. The dataset is then partitioned into training and testing sets.

A crucial part of the analysis involves computing CF, which serves as the target variable to measure boiler efficiency. Feature selection techniques are applied to identify key predictive variables, after which nonlinear regression models are developed, trained, and tested. Finally, prediction results are evaluated and compared to determine the model's accuracy and effectiveness, guiding improvements in operational efficiency for the boiler's convective components.

Dataset Preparation: Preprocessing and Partitioning Data

This dataset was extracted from a 1,000 MW ultra-supercritical coal-fired power plant in Malaysia, covering the years 2018 and 2023. Data from 2018 is considered the reference state of cleanliness for the boiler, while data from 2023 represent the current state of the boiler under investigation. According to plant engineers and technicians, the data from 2018 is considered to represent the cleanest operating conditions, as it marks the beginning of full-scale operations at the power plant. Before that, the plant was still in its trial phase. A total of 138 parameters were collected throughout the power plant. The raw data is primarily sourced from heat exchanger components to identify the parameters most

influencing boiler cleanliness. This process includes data preprocessing, CF computation, and data partitioning.

Preprocessing is essential to ensure model quality, as it helps eliminate outliers and manage missing values. Null data, where values were not recorded properly, is removed because most algorithms cannot process such entries. Outliers, defined as data with a target load below 750 MW, are also excluded. Data within the target load range is retained, as values below this threshold often correspond to the boiler's start-up phase before combustion stabilizes and are unsuitable for model development. After preprocessing, the dataset is randomly partitioned into training and testing sets, with an 80:20 ratio, where 80% of the data is used for training and 20% is reserved for testing.

Computing the Cleanliness Factor

In this study, the CF is selected as the dependent variable to assess fouling conditions in the convective section of the boiler. Equation 1 defines the calculation method for CF, which has been validated as an effective means of determining the cleanliness status of heat exchangers. CF is defined as the ratio of the real-time heat transfer rate, Q_r, to the heat transfer rate under clean conditions, Q_c (Breeding et al., 2010).

$$CF = \frac{Q_r}{Q_c} \qquad [1]$$

The equations for heat transfer rates under real-time and clean conditions ($Q_{r/c}$) are fundamentally similar, differing only by the timeframe of data collection. Reference data for the clean condition were recorded in 2018, marking the start of full-scale operations at the power plant. Therefore, this timeframe is considered to represent the boiler's cleanest state. Equation 2 defines the computation of the heat transfer rate, providing an accurate average based on the inlet and outlet temperatures of the gas and steam (Madejski et al., 2018).

$$Q_{r/c} = m\,(H_{outlet} - H_{inlet}) \qquad [2]$$

Here, m is the mass flow rate, and H_{outlet} and H_{inlet} represent the enthalpy values from the superheated steam table, based on the steam temperature and pressure within the boiler. This calculation ensures data consistency. Ultimately, this equation is adapted to align with the dataset, with CF designated as the target variable for training the nonlinear regression model.

Model Development

Model training in this study incorporates nonlinear regression models alongside feature selection techniques. Models with all 138 features represent those without feature selection,

as illustrated in the research model framework in Figure 2. Feature selection is performed using the MRMR algorithm, resulting in varying numbers of selected features. RT, SVR, ensembles of trees, and ANN are then developed based on the selected features.

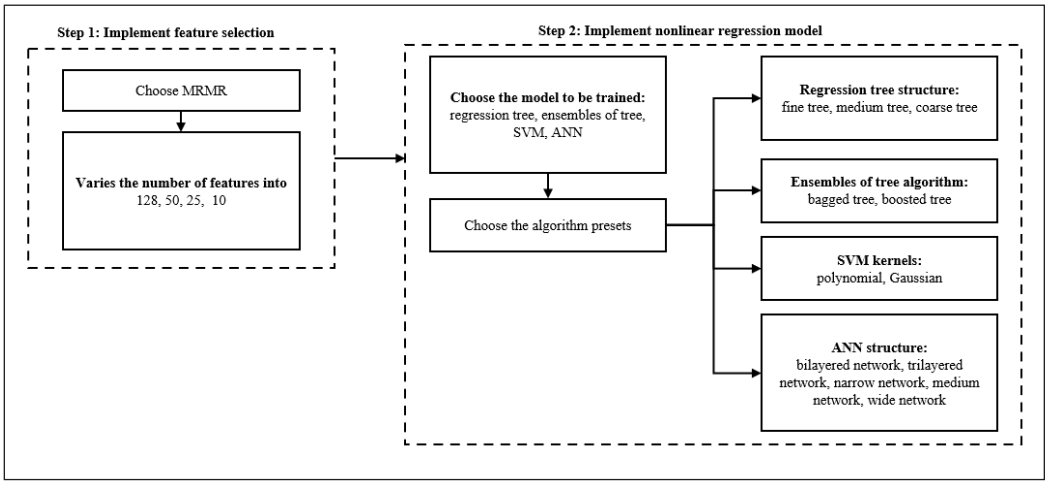


Figure 2. Structure of the research framework
Note. MRMR = Minimum redundancy maximum relevance; SVM = Support vector machine; ANN = Artificial neural network

Implementation of Feature Selection

The MRMR algorithm is a widely used feature selection method that identifies the most relevant features for predicting the target output while simultaneously minimizing redundancy among the selected features. By balancing relevance and redundancy, MRMR ensures that the chosen features provide highly informative and non-overlapping inputs to the model, ultimately enhancing the model's predictive accuracy and interpretability (Peng et al., 2005). This study adopts MRMR mainly due to its characteristics, which balance feature importance while still removing redundancy. Other techniques, such as recursive feature elimination (RFE), may have a high computational cost for large datasets (Ding et al., 2022) .

Prior to model training, feature selection is conducted using the MRMR algorithm. This step helps evaluate the impact of feature selection on model performance and stability, providing insight into the effectiveness of different feature sets. For each algorithm, feature subsets are selected in varying sizes—128, 50, 25, and 10 parameters, allowing the model to be trained with predetermined numbers of features. This approach enables a thorough assessment of the consistency and effectiveness of model performance based on different feature set sizes, supporting a robust evaluation of the role of feature selection in enhancing model reliability and generalizability.

Implementation of Nonlinear Regression Models

Models that correlate dependent and independent variables in a nonlinear manner to predict numerical outcomes effectively capture complex relationships, providing more accurate predictions of continuous data compared to linear models (Liang et al., 2022). The models selected for this study are RT, SVR, an ensemble of trees, and ANN. These models were chosen based on their suitability for handling nonlinear datasets, relatively low computational complexity, efficient processing time, and availability in the Matrix Laboratory (MATLAB). RT offers clear interpretability and can capture nonlinear relationships with minimal data preprocessing. SVR is particularly effective in high-dimensional datasets and maintains strong generalization performance due to its robustness against overfitting. An ensemble of trees combines multiple trees to enhance predictive performance, thereby improving accuracy, reducing overfitting, and increasing model robustness. Finally, ANN is widely applied in machine learning for its powerful capabilities in capturing complex and nonlinear patterns within data. It can also generalize well to unseen data. The following section provides a detailed discussion of the theory and function of each model.

RT

An RT is a decision tree used to predict continuous dependent variables. It functions by iteratively splitting the dataset into subsets based on the values of input features, aiming to partition the data into distinct regions. The process begins at the root node and continues through a series of splits until reaching the terminal leaves (Loh, 2011). Figure 3 illustrates the detailed structure of the RT splits. The tree structure shows how the CF is partitioned

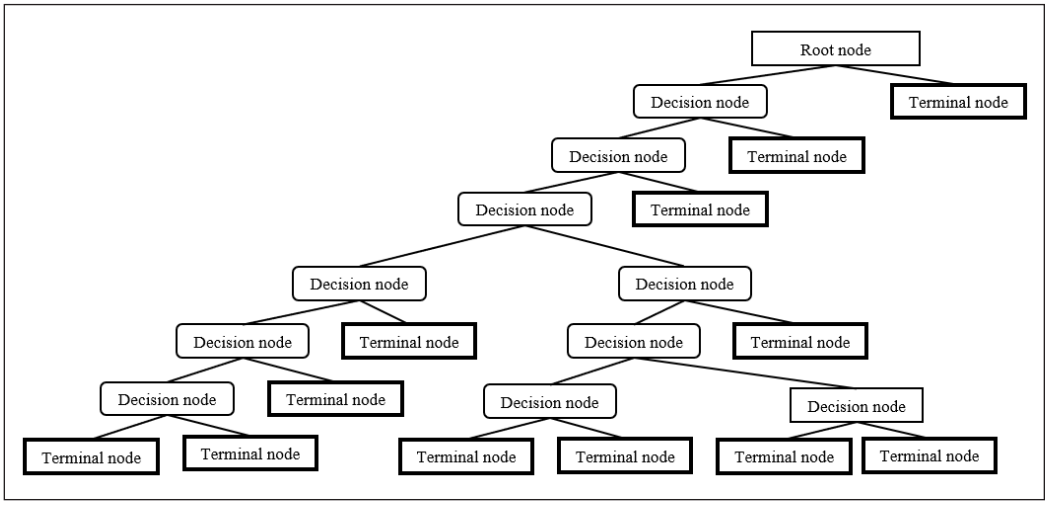


Figure 3. The regression tree structure with 12 terminal nodes

in Equation 3, where each data point (x_i) are segregated into the left and right sides based on the most relevant features (x_j) and their threshold (t), with the primary objective of minimizing prediction error.

The collected operational parameters serve as the input features, denoted as X, while the target values represent the outputs of the dataset.

$$X_{left} = \{x_i | x_j \leq t\}, X_{right} = \{x_i | x_j > t\} \quad [3]$$

The tree structure continues splitting nodes until a stopping criterion is met. Equation 4 represents the computation of predictions by applying new input features (x) within a specific region (R). The prediction, $y(x)$, is calculated as the average output over individual trees, expressed as the sum of outputs from each tree (y_i):

$$y(x) = \frac{1}{|R|} \sum y_i \quad [4]$$

There are three main types of RT structures based on the number of terminal nodes: fine, medium, and coarse trees. Fine trees create a structure with four terminal nodes, capturing complex patterns in the data but often resulting in overfitting, which can reduce the model's generalizability. Medium trees create a structure with 12 terminal nodes, providing a balance between complexity and generalization. Coarse trees, with 36 terminal nodes, are less prone to overfitting but may be less accurate in capturing patterns within the data (Cai et al., 2021).

Ensemble of Trees

Ensemble tree methods are regression algorithms that combine multiple decision tree models to achieve greater accuracy than a single predictive model (Ghiasi & Zendehboudi, 2021) . The methods applied in this paper are the bagged tree and the least-squares boosting tree (LSBoost). The bagged tree method employs bootstrap aggregation, also known as bagging, to enhance predictive performance and reduce variance by averaging the results of multiple decision trees. It creates an ensemble of trees by training multiple trees independently on random subsets of the training data. Equation 5 describes the final prediction $y(x)$, where the predictions from each individual tree, $y_m(x)$, are averaged. The random subsets are generated through bootstrap sampling, where samples are partitioned randomly with replacement, and M is the number of trees trained. Each tree is trained to minimize prediction error on its respective sample.

$$y(x) = \frac{1}{M} \sum_{m=1}^{M} y_m(x) \quad [5]$$

Boosted trees utilize the LSBoost algorithm to minimize the residual sum of squares between predicted and actual outputs, thereby enhancing model accuracy. This method is particularly suitable for weak learners because it improves the overall model performance. Each new model is trained sequentially to correct the errors made by previous models. While this algorithm reduces bias, it can be prone to overfitting.

Equation 6 represents the ensemble models $F_m(x)$, where $F_o(x)$ is the initial model that is calculated as the mean of the target values. The learning rate, denoted as v, partially influences the contribution of each weak learner. The function $h_m(x)$ represents the weak learner, which computes the difference between the actual CF and the current model predictions.

$$F_m(x) = F_0(x) + \sum_{m=1}^{M} v h_m(x) \qquad [6]$$

SVR

Support vector machines (SVM) have an extension known as SVR, which creates an optimal hyperplane based on the support vectors. Figure 4 illustrates the mechanism of SVR, where the hyperplane is a multidimensional surface that separates the data points, and the support vectors are the data points nearest to the hyperplane (Marafino et al., 2014). The goal of this algorithm is to maximize the margin that distinctly separates the data points. Equation 7 describes the function of parallel hyperplanes ($y(x)$).

$$f(\mathrm{x}) = w\mathrm{x} + b \qquad [7]$$

where w representing the margin and b is a constant.

The decision boundary is defined as $f(x) = 0$, which completely separates the two classes. Data points with $f(x) > 0$ belong to the red class, while data points with $f(x) < 0$ belong to the green class. Kernel functions are used to transform the data into a higher-dimensional space, aiming to model a nonlinear relationship between the input features and the CF. The kernels can be categorized into Gaussian SVM and polynomial SVM. The variations of kernels implemented in this study include quadratic, cubic, fine Gaussian, medium Gaussian, and coarse Gaussian.

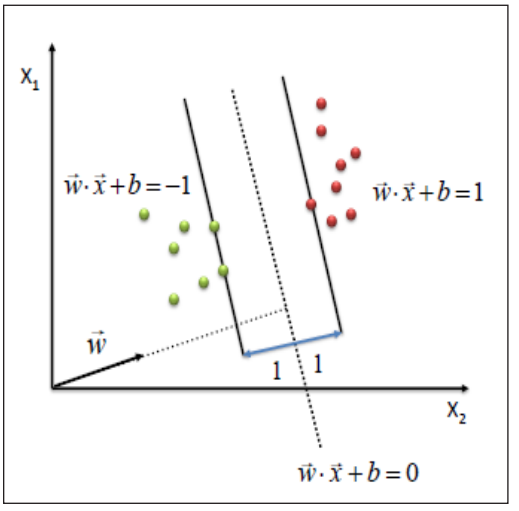


Figure 4. The illustration of the support vector regression mechanism (Marafino et al., 2014)
Note. $\vec{w}$ = Weight vector; $\vec{x}$ = Data point vector; b = Bias

ANN

An ANN is an algorithm inspired by the human brain's mechanisms, enabling it to handle non-linear relationships effectively. Figure 5 illustrates the architecture of a neural network, where the input layer contains the attributes of the input features, with each neuron representing a single feature. The output layer contains the neuron that holds the target value, which in this case is the CF values. In the case of a continuous target value or regression learner, the output layer consists of only one neuron. Between the input and output layers is the hidden layer, where the network learns to improve prediction accuracy. As the network expands, its capacity to learn complex relationships increases; however, this also raises the risk of overfitting. Backpropagation is employed to minimize the error between predicted and actual outputs (Comito & Pizzuti, 2022).

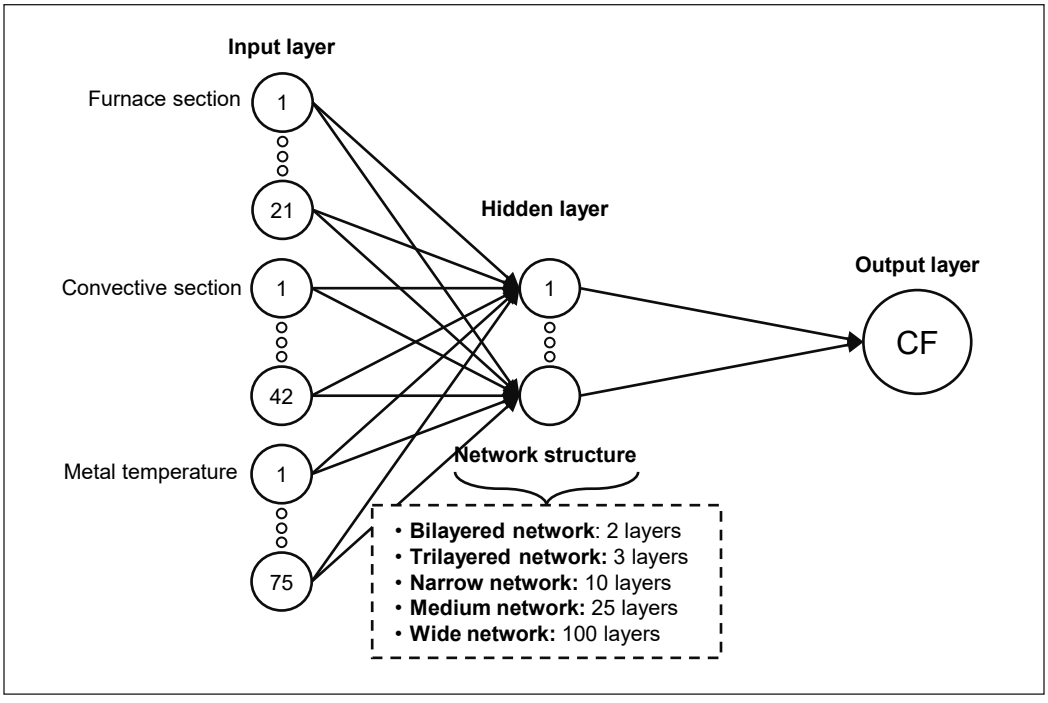


Figure 5. Artificial neural network
Note. CF = Cleanliness factor

Model Validation

After training the model, the prediction accuracy of the training and testing datasets has been analyzed using R^2 and MSE as key metrics. R^2 indicates the extent to which the variance in the dependent variable can be predicted based on the independent variable, with values closer to 1 indicating better model performance. Meanwhile, MSE measures

the average squared difference between predicted and actual values; thus, a smaller MSE indicates better performance. The equations for R^2 and MSE as shown in Equations 8 and 9.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad [8]$$

where y is the real plant data set or actual value, $\hat{y}$ is the prediction data set, and i is the number of data until n, the total number of samples.

$$R^2 = 1 - \frac{R_{RSS}}{R_{TSS}} \quad [9]$$

where R_{RSS} is the residual sum of squares between actual and predicted values and R_{TSS} is the sum of squared differences between actual values and the mean of actual values.

In addition, prediction speed and training time are also considered to provide insight into model efficiency. Prediction speed refers to the number of predictions the model can make per second, with higher values indicating better performance. Training time, on the other hand, is the amount of time required to train the model, with more complex models generally taking longer to train.

For this study, the analysis focuses on identifying the model with the highest R^2, lowest MSE, fewest selected features, highest prediction speed, and lowest training time. This criterion is established to find the best nonlinear regression models that are compatible with feature selection methods and capable of creating a robust nonlinear regression model.

RESULTS AND DISCUSSION

This section presents the simulation results and analyzes the performance of various regression models, including RT, SVR, ensembles of trees, and ANN. Each model is evaluated in detail, focusing on its strengths and weaknesses in predicting outcomes. The section concludes with a comparative analysis of nonlinear regression models, comparing the effectiveness of these models based on evaluation metrics.

Performance of the RT Model

The overall RT model demonstrates excellent performance on both datasets, as shown in Table 10. Across all trees, the MSE values remain consistently low, ranging from 0.001 to 0.002, indicating that the models fit the training data very well. The medium tree model exhibits the best performance, with R^2 values ranging from 0.973 to 0.958 for training and from 0.979 to 0.958 for testing.

In comparison, the fine tree model shows a gradual decline in performance, with R^2 values ranging from 0.973 to 0.953 for training and from 0.978 to 0.956 for testing. Lastly,

Table 10
Performance results of the regression tree

Preset	No. of features	MSE Training)	R^2 (Training)	MSE (Test)	R^2 (Test)
Fine tree	138	0.001	0.972	0.001	0.975
Fine tree	128	0.001	0.973	0.001	0.975
Fine tree	50	0.001	0.971	0.001	0.978
Fine tree	25	0.001	0.970	0.001	0.978
Fine tree	10	0.001	0.957	0.001	0.956
Medium tree	138	0.001	0.973	0.001	0.979
Medium tree	128	0.001	0.973	0.001	0.979
Medium tree	50	0.001	0.971	0.001	0.976
Medium tree	25	0.001	0.972	0.001	0.976
Medium tree	10	0.001	0.958	0.001	0.958
Coarse tree	138	0.001	0.969	0.001	0.975
Coarse tree	128	0.001	0.969	0.001	0.976
Coarse tree	50	0.001	0.968	0.001	0.968
Coarse tree	25	0.001	0.967	0.001	0.973
Coarse tree	10	0.002	0.953	0.002	0.954

Note. MSE = Mean squared error

the coarse tree model delivers the lowest performance, with R^2 values ranging from 0.969 to 0.953 for training and from 0.975 to 0.954 for testing.

The worst-performing model is the coarse tree with 10 selected features, achieving R^2 values of 0.953 for training and 0.954 for testing, alongside MSE values of 0.002 for both datasets. In contrast, the medium tree model with 10 features shows significantly better performance, attaining an R^2 value of 0.958 and an MSE value of 0.001 for both datasets. This indicates that the medium tree model achieves optimal performance while maintaining computational efficiency with a smaller subset of features.

Performance of the Ensemble of Trees

The simulation results presented in Table 11 indicate that the bagged trees exhibit exceptional performance on both datasets, with R^2 values ranging from 0.973 to 0.983 for training and from 0.976 to 0.990 for testing, along with MSE values of 0.001 for training and between 0 and 0.001 for testing. This performance suggests that the model is highly effective at fitting the dataset with minimal variation across feature subsets.

In comparison, the performance of the boosted trees is lower, with R^2 values ranging from 0.874 to 0.891 for training and from 0.872 to 0.899 for testing, as well as MSE values of 0.004 for training and between 0.003 and 0.004 for testing. This indicates that the boosted tree is less effective for this specific dataset, although it still performs adequately.

Table 11
Performance results of the ensemble of trees

Preset	No. of features	MSE (Training)	R^2 (Training)	MSE (Test)	R^2 (Test)
Bagged tree	138	0.001	0.983	0	0.990
Bagged tree	128	0.001	0.983	0	0.990
Bagged tree	50	0.001	0.982	0	0.989
Bagged tree	25	0.001	0.981	0	0.988
Bagged tree	10	0.001	0.973	0.001	0.976
Boosted tree	138	0.004	0.891	0.003	0.899
Boosted tree	128	0.004	0.891	0.003	0.898
Boosted tree	50	0.004	0.89	0.003	0.896
Boosted tree	25	0.004	0.887	0.004	0.893
Boosted tree	10	0.004	0.874	0.004	0.872

Note. MSE = Mean squared error

Overall, the boosted tree model with 10 features shows the worst performance, achieving R^2 values of 0.874 for training and 0.872 for testing, along with an MSE value of 0.004 for both datasets. In contrast, the bagged trees model with 10 features demonstrates excellent performance, achieving R^2 values of 0.973 for training and 0.976 for testing, with an MSE value of 0.001 for both datasets.

Performance of the SVR

For this study, two types of kernels are used: Gaussian and polynomial kernels. The Gaussian kernels consist of fine Gaussian, medium Gaussian, and coarse Gaussian, while the polynomial kernels comprise quadratic and cubic kernels. Based on the results documented in Table 12, Gaussian kernels demonstrate better performance compared to polynomial kernels. This is evident from the R^2 values ranging from 0.798 to 0.990 for training and from 0.785 to 0.982 for testing, along with MSE values ranging from 0.001 to 0.008 on both datasets.

In comparison, polynomial kernels exhibit a drastic performance reduction, with R^2 values ranging from -236.941 to 0.898 for training and from -35.709 to 0.966 for testing, alongside MSE values ranging from 0.003 to 56.522 for training and from 0.001 to 1.282 for testing. Despite the polynomial kernels showing better performance within the test datasets, the significant differences between the training and testing results suggest potential issues of overfitting or underfitting when feature selection is applied.

Among Gaussian kernels, the medium Gaussian kernel shows excellent performance, with R^2 values ranging from 0.965 to 0.985 for training and from 0.966 to 0.938 for testing, along with MSE values ranging from 0 to 0.001 for training and from 0.001 to 0.002 for testing. Following that, the coarse Gaussian kernel achieves R^2 values ranging

Table 12
Performance results of the support vector regression

Preset	No. of features	MSE (Training)	R^2 (Training)	MSE (Test)	R^2 (Test)
Fine Gaussian	138	0.007	0.798	0.008	0.785
Fine Gaussian	128	0.006	0.836	0.006	0.823
Fine Gaussian	50	0.004	0.878	0.005	0.867
Fine Gaussian	25	0.004	0.880	0.004	0.873
Fine Gaussian	10	0.004	0.893	0.004	0.887
Medium Gaussian	138	0	0.989	0.001	0.966
Medium Gaussian	128	0	0.990	0.001	0.964
Medium Gaussian	50	0	0.990	0.001	0.958
Medium Gaussian	25	0.001	0.984	0.002	0.951
Medium Gaussian	10	0.001	0.970	0.002	0.938
Coarse Gaussian	138	0.001	0.979	0.001	0.982
Coarse Gaussian	128	0.001	0.980	0.001	0.982
Coarse Gaussian	50	0.001	0.979	0.001	0.982
Coarse Gaussian	25	0.001	0.971	0.001	0.970
Coarse Gaussian	10	0.002	0.943	0.002	0.937
Quadratic	138	0.007	0.795	0.001	0.966
Quadratic	128	0.128	-2.781	0.002	0.950
Quadratic	50	0.003	0.898	0.007	0.787
Quadratic	25	0.198	-4.838	0.022	0.374
Quadratic	10	0.006	0.835	0.053	-0.527
Cubic	138	0.185	-4.451	0.048	-0.364
Cubic	128	6.109	-178.741	0.053	-0.508
Cubic	50	56.522	-1662.100	0.067	-0.919
Cubic	25	8.087	-236.941	0.015	0.570
Cubic	10	2.706	-78.625	1.282	-35.709

Note. MSE = Mean squared error

from 0.980 to 0.943 for training and from 0.982 to 0.937 for testing, with MSE values ranging from 0.001 to 0.002 for both datasets. Lastly, the fine Gaussian kernel delivers lower performance, with R^2 values ranging from 0.798 to 0.893 for training and from 0.785 to 0.887 for testing, alongside MSE values ranging from 0.004 to 0.008 for both datasets.

The significance of feature selection in improving model accuracy is evident throughout the evaluations, particularly in the SVR models with fine Gaussian kernels. The models

incorporating feature selection consistently achieve higher R^2 values, indicating a better fit and enhanced predictive performance. These findings highlight the critical role of feature selection in boosting the efficiency and reliability of machine learning models.

The worst-performing model is the cubic kernel with 25 features selected, achieving R^2.values of -236.941 for training and 0.570 for testing, and MSE values of 8.087 for training and 0.015 for testing. In comparison, the best-performing model is the medium Gaussian kernel with 10 features selected, achieving R^2 values of 0.970 for training and 0.938 for testing, with MSE values of 0.001 for training and 0.002 for testing.

Performance of the ANN

Overall, all network structures perform exceptionally well compared to other algorithms, as shown in Table 13. This is evident with R^2 values ranging from 0.91 to 0.99 for training, 0.817 to 0.993 for testing, and MSE values ranging from 0 to 0.006 in both datasets. Among the various network structures, it is particularly challenging to select the most suitable one. Despite this, the worst-performing network structure is the trilayered network, which has R^2 values ranging from 0.91 to 0.986 for training, 0.817 to 0.984 for testing, and MSE values ranging from 0 to 0.006 in both datasets.

The most accurate models for each algorithm, regardless of the number of feature subsets, have been identified. The bilayered network with 128 selected features achieves R^2 values of 0.987 for training and 0.966 for testing, with MSE values of 0 for training and 0.001 for testing. The trilayered network with 128 selected features achieves R^2 values of 0.987 for training and 0.817 for testing, with MSE values of 0 for training and 0.006 for testing. The narrow network with 128 features achieves R^2 values of 0.989 for training and 0.988 for testing, with MSE values of 0 for both datasets. The medium network with 128 features achieves R^2 values of 0.988 for training and 0.993 for testing, along with MSE values of 0 for both datasets as well. Lastly, the wide network with 128 features achieves R^2 values of 0.99 for training and 0.993 for testing, with MSE values of 0 for both datasets.

Despite the models' superior accuracy, they are not considered the optimal choice due to their reliance on numerous variables, which increases the risk of overfitting and reduces the models' generalizability. Overall, all network structures exhibit a great balance in performance, but their effectiveness gradually declines as the number of features is reduced to 10.

The worst-performing model is the trilayered model with 10 selected features, achieving R^2 values of 0.91 for training and 0.978 for testing, along with MSE values of 0.003 for training and 0.001 for testing. In comparison, the best-performing model is the medium neural network with 10 selected features, achieving R^2 values of 0.961 for training and 0.988 for testing, with MSE values of 0.001 for both datasets.

Table 13
Performance results of the artificial neural network

Preset	No. of features	MSE (Training)	R^2 (Training)	MSE (Test)	R^2 (Test)
Bilayered	138	0.001	0.979	0.004	0.884
Bilayered	128	0	0.987	0.001	0.966
Bilayered	50	0.001	0.979	0	0.991
Bilayered	25	0.001	0.967	0.001	0.972
Bilayered	10	0.002	0.944	0.001	0.984
Trilayered	138	0	0.986	0.001	0.984
Trilayered	128	0	0.987	0.006	0.817
Trilayered	50	0.001	0.984	0.003	0.918
Trilayered	25	0.001	0.973	0.001	0.971
Trilayered	10	0.003	0.910	0.001	0.978
Narrow	138	0.001	0.983	0.001	0.959
Narrow	128	0	0.989	0	0.988
Narrow	50	0.001	0.978	0.001	0.970
Narrow	25	0.001	0.973	0.001	0.969
Narrow	10	0.001	0.957	0.001	0.978
Medium	138	0	0.987	0.001	0.964
Medium	128	0	0.988	0	0.993
Medium	50	0.001	0.982	0.001	0.978
Medium	25	0.001	0.967	0.001	0.984
Medium	10	0.001	0.961	0.001	0.980
Wide	138	0	0.986	0.001	0.959
Wide	128	0	0.990	0	0.993
Wide	50	0.001	0.985	0	0.992
Wide	25	0.001	0.976	0	0.987
Wide	10	0.002	0.946	0.001	0.975

Note. MSE = Mean squared error

Comparative Analysis of Nonlinear Regression Models

The best model from each algorithm has been selected for evaluation, as summarized in Table 14 for detailed comparison and analysis. The final training step compared model performance based on the highest R^2 value, lowest MSE, and fewest selected features, providing a robust measure of each model's predictive capability in relation to feature selection.

The bagged tree model performed exceptionally well in terms of overall accuracy, achieving the highest R^2 value of 0.973 (training) and 0.976 (test), along with low MSE values of 0.001 for both datasets. In Figure 6, there are a few outliers in the high-value

Table 14
Summary of the best-performing model

Model	No. of features	MSE (Training)	R^2 (Training)	MSE (Test)	R^2 (Test)
Bagged tree	10	0.001	0.973	0.001	0.976
Medium NN	10	0.001	0.961	0.001	0.980
Medium Gaussian SVR	10	0.001	0.970	0.002	0.938
Medium tree	10	0.001	0.958	0.001	0.958

Note. MSE = Mean squared error; NN = Neural network; SVR = Support vector regression

range, but they are minimal, indicating that the bagged tree model provides highly accurate predictions with only a small margin of error. The regression line equation of this model is $y = 0.9675x + 0.0378$, and achieves the highest R^2 value among all, indicating that the predicted values explain 97.3% of the variance in the actual values. The slope of 0.9675, which is very close to 1, implies that the prediction scales well with the actual values. Meanwhile, the small positive interception of 0.03781 may have introduced a very slight upward bias. Overall, the bagged tree model achieves high accuracy and low bias, making it the most reliable model in this comparison.

The second-best performer is the medium neural network (NN), with R^2 values of 0.961 (training) and 0.98 (test), and low MSE values of 0.001 for both datasets. In Figure 7, a few outliers appear around value 6, which are much larger than the true response, though most predictions cluster near the diagonal line. This suggests that the medium NN model demonstrates strong predictive capability but may handle more outliers than the bagged tree model. The regression line equation of this model is $y = 0.9874x + 0.01478$, the lower variance of 96.1% compared to the bagged tree shows it is slightly less reliable. Its slope is closest to 1, and the interception is minimal. However, a drop in R^2 suggests slightly more errors in generalization and prediction variability.

The third performer is the medium Gaussian SVR, with R^2 values of 0.97 (training) and 0.938 (test), and MSE values of 0.001 (training) and 0.002 (test). In Figure 8, predictions generally scatter near the diagonal line but not as closely, indicating that the medium Gaussian SVR model has good prediction accuracy but may be more prone to error. The regression line equation of this model is $y = 0.9637x + 0.0418$, with a strong R^2 value is nearly on par with the bagged tree. The slope is a little lower with a value of 0.9637, meaning predictions increase slightly less rapidly compared to actual values. The slightly higher interception suggests a small, consistent overestimation. While it is a strong model, its slightly lower slope and higher intercept are showing an inferior performance as compared to the bagged tree model.

The final performer is the medium tree, with R^2 values of 0.958 for both datasets, and MSE values of 0.001 for both datasets. This performance comparison suggests that these

models effectively reduce computational complexity while maintaining strong predictive accuracy and efficiency. In Figure 9, more outliers appear compared to other models, suggesting that the fine tree model struggles with outliers more than the others, leading to increased errors. The regression line equation of this model is $y = 0.9667x + 0.03852$, with the lowest R^2 value among the four models, suggesting it is less effective at capturing the variability of the target variable. The slope and intercept are closely similar to the bagged tree, but the reduced R^2 value implies more prediction error and weaker model generalization. This model performs decently well but is the least preferable among the four.

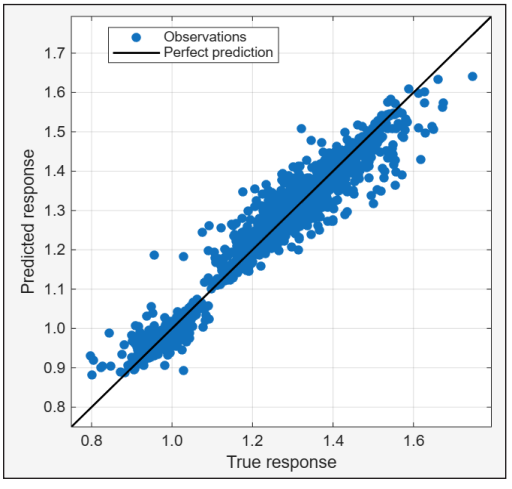


Figure 6. Response of bagged trees with 10 selected features

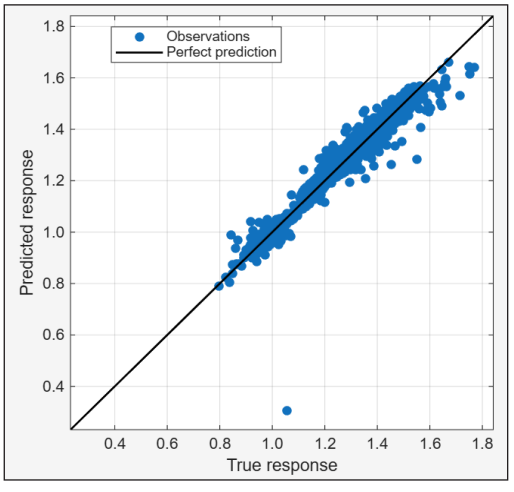


Figure 7. Response of the medium artificial neural network with 10 selected features

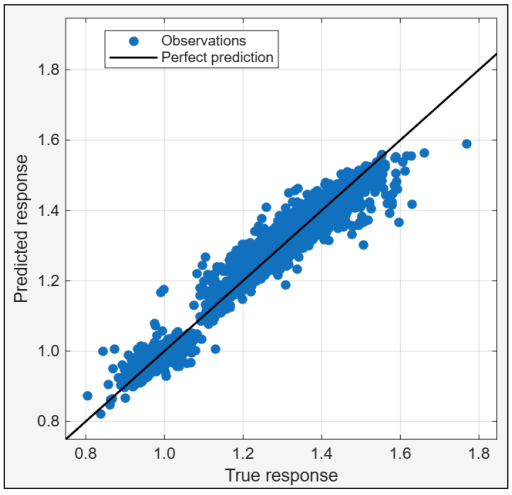


Figure 8. Response of the medium Gaussian support vector regression with 10 selected features

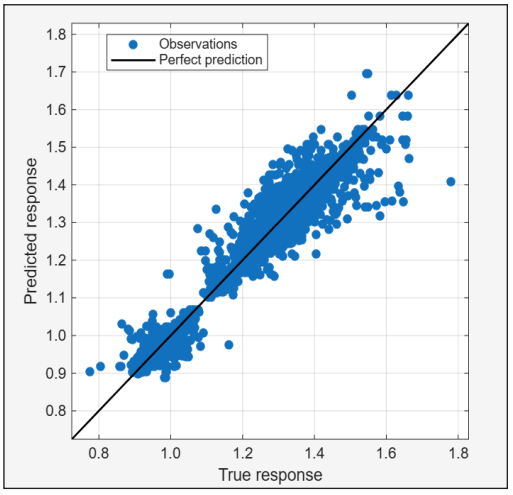


Figure 9. Response of the medium tree with 10 selected features

Overall, the bagged tree model is designated as the best model due to its balance of high predictive accuracy, efficient use of a smaller, well-selected feature set, and strong capability to minimize deviations caused by outliers. Moreover, the bagged tree model consistently demonstrates excellent performance across both datasets, even with smaller feature subsets.

CONCLUSION

This research effectively analyzed nonlinear regression models to predict the CF within the convective sections of the boiler, highlighting the models' practicality and performance in these crucial areas. The study identified bagged trees, a medium neural network, a medium Gaussian SVR, and a medium tree, all with 10 features selected to be the top-performing machine learning models that required fewer feature subsets, which is essential for optimizing boiler performance. For example, the bagged trees achieve an R^2 of 0.99 without feature selection and 0.976 with only 10 selected features, resulting in a 93% reduction from the original input. This reduction not only enhances computational speed and reduces memory usage but also improves model interpretability and operational feasibility.

Comparatively, while NNs and SVR provided slightly higher accuracy, they demanded more computational resources, especially in the training and tuning phases. On the other hand, tree-based models such as bagged trees and medium trees offered a balanced trade-off between performance and processing efficiency, allowing them to be more suitable for real-time deployments. Moreover, implementing these models in operational environments exposes them to challenges such as real-time sensor data, managing model complexity during control system integration, and maintaining operational parameter settings. Therefore, careful consideration must be given to sensor calibration and data preprocessing pipelines to ensure robust integration.

Future research should focus on integrating reinforcement learning and online monitoring machine learning techniques to adjust predictions based on live operational data dynamically. By pinpointing these features more accurately, the study can support the development of a more targeted and efficient soot-blowing mechanism. This would enhance operational precision, leading to improved boiler efficiency, cost savings, and extended equipment lifespan.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Higher Education under the Fundamental Research Grant Scheme (FRGS/1/2024/TK07/UPM/02/3). Special thanks are extended to Sultan Azlan Shah Power Station, Tenaga Nasional Berhad, for providing data and consultation for this project.

REFERENCES

Bezerra, F. E., de Oliveira Neto, G. C., Cervi, G. M., Mazetto, R. F., de Faria, A. M., Vido, M., Lima, G. A., de Araújo, S. A., Sampaio, M., & Amorim, M. (2024). Impacts of feature selection on predicting machine failures by machine learning algorithms. *Applied Sciences*, *14*(8), 3337. https://doi.org/10.3390/app14083337

Breeding, C., Tandra, D., & Shah, S. (2010). Boiler cleaning using ISB (Intelligent Soot Blowing) system integration: Recent developments and case study. In *ASME 2010 Power Conference* (pp. 89–96). American Society of Mechanical Engineering. https://doi.org/10.1115/power2010-27322

Cai, H., Qin, W., Wang, L., Hu, B., & Zhang, M. (2021). Hourly clear-sky solar irradiance estimation in China: Model review and validations. *Solar Energy*, *226*, 468–482. https://doi.org/10.1016/j.solener.2021.08.066

Comito, C., & Pizzuti, C. (2022). Artificial intelligence for forecasting and diagnosing COVID-19 pandemic: A focused review. *Artificial Intelligence in Medicine*, *128*, 102286. https://doi.org/10.1016/j.artmed.2022.102286

Ding, X., Yang, F., & Ma, F. (2022). An efficient model selection for linear discriminant function-based recursive feature elimination. *Journal of Biomedical Informatics*, *129*, 104070. https://doi.org/10.1016/j.jbi.2022.104070

Ghiasi, M. M., & Zendehboudi, S. (2021). Application of decision tree-based ensemble learning in the classification of breast cancer. *Computers in Biology and Medicine*, *128*, 104089. https://doi.org/10.1016/j.compbiomed.2020.104089

Jemai, J., & Zarrad, A. (2023). Feature selection engineering for credit risk assessment in retail banking. *Information*, *14*(3), 200. https://doi.org/10.3390/info14030200

Kumari, S. A., & Srinivasan, S. (2019). Ash fouling monitoring and soot-blow optimization for reheater in thermal power plant. *Applied Thermal Engineering*, *149*, 62–72. https://doi.org/10.1016/j.applthermaleng.2018.12.031

Li, P., Li, K., Zhou, Y., Li, Q., Shi, Z., & Zhong, W. (2023). An effective strategy for monitoring slagging location and severity on the waterwall surface in operation coal-fired boilers. *Energies*, *16*(24), 7925. https://doi.org/10.3390/en16247925

Li, Q., Yan, P., Liu, J., Shi, Y., & Yu, D. (2020). Prediction of pollution state of heating surface in coal-fired utility boilers. *IEEE Access*, *8*, 206132–206145. https://doi.org/10.1109/access.2020.3036840

Liang, D., Frederick, D. A., Lledo, E. E., Rosenfield, N., Berardi, V., Linstead, E., & Maoz, U. (2022). Examining the utility of nonlinear machine learning approaches versus linear regression for predicting body image outcomes: The U.S. Body Project I. *Body Image*, *41*, 32–45. https://doi.org/10.1016/j.bodyim.2022.01.013

Loh, W.-Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery*, *1*(1), 14–23. https://doi.org/10.1002/widm.8

Madejski, P., Janda, T., Taler, J., Nabagło, D., Węzik, R., & Mazur, M. (2018). Analysis of fouling degree of individual heating surfaces in a pulverized coal fired boiler. *Journal of Energy Resources Technology*, *140*(3), 032003. https://doi.org/10.1115/1.4037936

Marafino, B. J., Davies, J. M., Bardach, N. S., Dean, M. L., & Dudley, R. A. (2014). N-gram support vector machines for scalable procedure and diagnosis classification, with applications to clinical free text data from the intensive care unit. *Journal of the American Medical Informatics Association*, *21*(5), 871–875. https://doi.org/10.1136/amiajnl-2014-002694

Menn, N., & Chudnovsky, B. (2021). FTR-based expert system for power generation units. In A. J. Tallón-Ballesteros (Ed.), *Modern management based on Big Data II and Machine Learning and Intelligent Systems III* (Vol. 341, pp. 548-556). IOP Press. https://doi.org/10.3233/faia210287

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(8), 1226–1238. https://doi.org/10.1109/tpami.2005.159

Shi, Y., Zhang, Z., Li, M., Wen, J., & Cui, F. (2021). Soot blowing optimization for coal-fired boilers with ash accelerated deposition model based on gamma process. In *40th Chinese Control Conference* (pp. 6617-6622). IEEE. https://doi.org/10.23919/ccc52363.2021.9549924

Shi, Y., Li, M., Wang, J., Wen, J., Cui, F., & Qiao, G. (2022). Prediction of ash deposition on heating surfaces of coal-fired power plant boiler based on dynamic neural network. In *34th Chinese Control and Decision Conference* (pp. 779–783). IEEE. https://doi.org/10.1109/ccdc55256.2022.10034385

Tenaga Nasional Berhad. (2024). *Ultra-supercritical boiler power plant*. TNB.

Thota, S., & Syed, M. B. (2024). Analysis of feature selection techniques for prediction of boiler efficiency in case of coal based power plant using real time data. *International Journal of Systems Assurance Engineering and Management*, *15*, 300–313. https://doi.org/10.1007/s13198-022-01725-y

Wei, W., Cheng, S., & Sun, F. (2020). Research on formation mechanism of typical low-temperature fouling layers in coal-fired boilers. *Fuel*, *261*, 116215. https://doi.org/10.1016/j.fuel.2019.116215

Wen, J., Shi, Y., Pang, X., & Jia, J. (2022). Optimal soot blowing and repair plan for boiler based on HJB equation. *Optimization*, *71*(16), 4603–4622. https://doi.org/10.1080/02331934.2021.1954922

Pertanika J. Sci. & Technol. 33 (6): 2747 - 2770 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Development and Verification of an MCNP5-based Gamma Camera for ^{177}Lu SPECT Imaging

Puvanesuawary Morthy[1], Marianie Musarudin[1]*, Nor Shazleen Ab Shukor[1], Mohamad Aminudin Said[2], Dong Xianling[3] and M. Iqbal Saripan[4]

[1]School of Health Sciences, Health Campus, Universiti Sains Malaysia, 16150 Kota Bharu, Kelantan, Malaysia
[2]Department of Nuclear Medicine, Institut Kanser Negara, 62250 Putrajaya, Wilayah Persekutuan Putrajaya, Malaysia
[3]Hebei International Research Center of Medical Engineering, Chengde Medical University, Chengde 067050, China
[4]Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

ABSTRACT

Monte Carlo (MC) simulations in single-photon emission computed tomography (SPECT) are invaluable for enhancing activity quantification in patients, thereby facilitating an accurate estimate of the absorbed dose for every organ. The validation of MC results against experimental data is essential to ensure their reliability. This study focuses on lutetium-177 (^{177}Lu) SPECT imaging modeling using the Monte Carlo N-Particle Transport, version 5 (MCNP5) code. The objective was to validate the MC code for ^{177}Lu SPECT by comparing simulated and experimental data on system sensitivity, image quality, intensity profiles, and energy resolution. The gamma camera was modeled based on the GE HealthCare Discovery nuclear medicine/computed tomography (NM/CT) 670 Pro, with a medium energy general purpose (MEGP) collimator. MC simulations accounted for the SPECT scanner's physical specifications, Gaussian energy blurring, and corrections for scatter and attenuation. Post-simulation analysis was performed in the Matrix Laboratory (MATLAB) software, with comparisons made under identical geometrical configurations for experimental and simulated data. The results revealed a strong agreement between simulated and experimental data. System sensitivities varied by only 3.34%, while the structural similarity index metric (SSIM) value of 0.9348 demonstrated a high level of similarity. The energy resolutions for 113 and 208 keV were 16.28 and 15.69%, respectively. Additionally, high correlation coefficients across spheres further validated the accuracy of the MC model. These remarkable agreements confirm the accuracy and precision of the MCNP5-based

ARTICLE INFO

Article history:
Received: 28 March 2025
Accepted: 16 July 2025
Published: 14 October 2025

DOI: https://doi.org/10.47836/pjst.33.6.15

E-mail addresses:
puvanesmorthy0698@gmail.com (Puvanesuawary Morthy)
marianie@usm.my (Marianie Musarudin)
shazzleen@usm.my (Nor Shazleen Ab Shukor)
aminhpj@gmail.com (Mohamad Aminudin Said)
dongxl_cdmc@163.com (Dong Xianling)
iqbal@upm.edu.my (M. Iqbal Saripan)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

^{177}Lu SPECT imaging simulations. In conclusion, the validated MCNP5 model of the gamma camera provides a foundation for future research into ^{177}Lu SPECT imaging and quantification.

Keywords: ^{177}Lu, Monte Carlo, SPECT, verification

INTRODUCTION

SPECT is now widely recognized as a key tool in nuclear medicine research for the localization of *in vivo* radiopharmaceuticals, as it can generate three-dimensional images (Bouchareb et al., 2024). Through detecting gamma photons emitted from radiopharmaceuticals and their reconstruction into tomographic images, SPECT offers crucial insights into the physiological functions of organs and tissues that other imaging modalities often do not provide (Mishra, 2024). This modality is widely utilized in oncology, cardiology, and neurology, where an accurate quantitative assessment of radioactivity in tissues is vital for effective diagnosis, treatment, and dosimetry evaluation. The combination of SPECT with computed tomography (CT) has led to the development of non-invasive hybrid SPECT/CT clinical systems, which have improved clinical applications by integrating anatomical (CT) and functional (SPECT) imaging, resulting in widespread adoption globally since 1999 (Zaidi, 1999). These systems play a crucial role in cancer diagnosis and treatment, enhancing the accuracy and precision of radiotherapy planning (Seo et al., 2008). However, the effectiveness of SPECT imaging is heavily dependent on the performance of the gamma camera.

In recent years, MC simulations have become increasingly important for assessing and optimizing gamma camera performance. MC models are frequently employed across various nuclear medicine modalities to refine imaging protocols, evaluate collimator designs, and improve quantitative accuracy (Dong et al., 2018; Musarudin et al., 2015; Ramonaheng et al., 2020; Saripan et al., 2009). Validating an MC model is mandatory before using it to simulate a SPECT imaging system. Due to the approximations and simplifications inherent in the physics laws formulated within an MC code (Di Domenico et al., 2023), validation is important to ensure accuracy between simulated and experimental measurements across key parameters. This comparison not only identifies discrepancies but also helps refine the MC model for improved accuracy.

Various MC codes, including MCNP, simulating medical imaging nuclear detectors (SIMIND) and GEometry ANd Tracking 4 (GEANT4), have been extensively utilized to model gamma cameras for SPECT imaging, with numerous studies confirming their accuracy and reliability. For instance, Bahnamiri et al. (2015) used MCNP calculations to explore the effect of collimator parameters on SPECT image quality and discovered a relative difference of less than 5% between simulation and experimental data, validating the accuracy of the simulation. Recently, Di Domenico et al. (2023) demonstrated that the

SIMIND MC code for gamma camera modeling is effective in simulating both planar and SPECT gamma camera systems for optimal quantification by validating calibration factors (CF) and recovery coefficients (RC) derived using this code. Similarly, Pells et al. (2023) explored simulation and experimental uncertainties in the spatial resolution, sensitivity, energy spectra, and projection images using the GEANT4 Application for Tomographic Emission (GATE) toolkit. Bui et al. (2023) evaluated the performance of the GEANT4-based MC model of a SPECT scanner in terms of detection efficiency, spatial resolution, and activity RC, achieving good agreement between the simulated and experimental data. Collectively, these studies demonstrate the ability of MC simulations to accurately model key parameters, such as the collimator influence, CF, RC, spatial resolution, and detection efficiency. The close match between the simulated and experimental data highlights the potential of MC models to improve SPECT quantification and image quality, emphasizing their role in advancing the accuracy and effectiveness of SPECT imaging systems.

Researchers typically validate simulations of nuclear imaging modalities by examining various parameters, including the spatial resolution, planar sensitivity, CF, and RC for isotopes like technetium-99m (^{99m}Tc), ^{177}Lu, and iodine-131 (^{131}I) (Bahreyni Toossi et al., 2010; Dong et al., 2018; Ejeh et al., 2019; Ramonaheng et al., 2020). These parameters offer valuable insights into the performance of MC codes in modeling gamma cameras, ensuring their effectiveness for future clinical applications (Morphis et al., 2021). ^{177}Lu has emerged as a promising radionuclide for targeted radionuclide therapy, namely in the treatment of neuroendocrine tumors and prostate cancer (Bodei et al., 2011; Ezziddin et al., 2014; Ilan et al., 2015; Paganelli et al., 2014; Romer et al., 2014; van der Zwan et al., 2015). With its favorable decay characteristics, beta emissions for therapy, and gamma emissions at 113 and 208 keV for imaging, ^{177}Lu has become a key radionuclide in theranostic applications, highlighting the necessity for precise SPECT quantification of ^{177}Lu in patient studies.

While MC simulations hold promise for optimizing SPECT imaging, accurately modeling complex clinical systems for quantitative analysis remains challenging. This study aims to develop and evaluate an MC-modeled gamma camera for Discovery NM/CT 670 Pro using the MCNP5 with ^{177}Lu. In our previous study, a simulation framework based on MNCP5 was successfully introduced to derive CFs and RCs for ^{177}Lu quantification, but it only offered a preliminary verification of the gamma camera model (Morthy et al., 2025). A comprehensive, system-wide validation that is essential for establishing the accuracy and precision of the model was not thoroughly explored. Notably, key performance metrics such as qualitative and quantitative image quality, sphere-wise intensity profiles, full width at half maximum (FWHM) analysis, and energy resolution were not systematically assessed. The current study addresses these gaps by performing a rigorous, multidimensional validation of an MCNP5 model meticulously tailored to the GE Discovery NM/CT 670 Pro gamma camera equipped with a MEGP collimator for ^{177}Lu SPECT imaging. This validation

includes thorough comparisons with experimental data to ensure high spatial, structural, and energy fidelity in the simulation. The enhanced agreement between simulation and experimental data significantly improves the MC model's reliability for future applications in personalized dosimetry, protocol optimization, and quantitative SPECT research. By enhancing MCNP5 for SPECT scanner simulation, this research seeks to improve imaging protocols and aid the development of innovative reconstruction algorithms.

MATERIALS AND METHODS

This research was divided into three phases: system sensitivity measurement, National Electrical Manufacturers Association (NEMA) phantom SPECT imaging and MC simulation, and verification of the MC-modeled gamma camera. The first phase involved assessing the system's ability to detect the incident photons and establishing baseline sensitivity data. In the second phase, a NEMA phantom was scanned using a SPECT system, while a corresponding MC simulation was run to model the SPECT imaging process. The acquisition and reconstruction settings determined using the GE Discovery NM/CT 670 Pro SPECT system in the previous study by Morthy (2025) were applied to both the experimental and simulation SPECT systems in this study. A compromise between the optimal image quality and quantification accuracy was achieved by employing the MEGP collimator, a slice thickness of 5 mm, and a reconstruction algorithm employing ordered subset expectation maximization (OSEM), where two iterations and 10 subsets constituted the optimal approach (Morthy, 2025). Dead time effects were considered negligible for the data acquired in this study, as they were estimated to be only a few percent based on previous studies (Nakanishi et al., 2024). The third phase aimed to verify the accuracy of the MC-modeled gamma camera by comparing simulated and experimental data. A 10% uncertainty was considered sufficient for the objectives of this study, with the simulation likely converging with the targeted 68% confidence interval. Table 1 summarizes key characteristics of ^{177}Lu, including its half-life, primary gamma radiation emissions, and maximum energy of its beta radiation emissions.

Table 1
Physical properties of a lutetium-177 (^{177}Lu) radioisotope

Radionuclide	Half-life ($t_{1/2}$, days)	Strongest γ emission Eγ (keV) (Iγ [%])	Maximum β energy (Emax, keV)
^{177}Lu	6.65	112.9 (6.2) 208.4 (10.4)	498.3

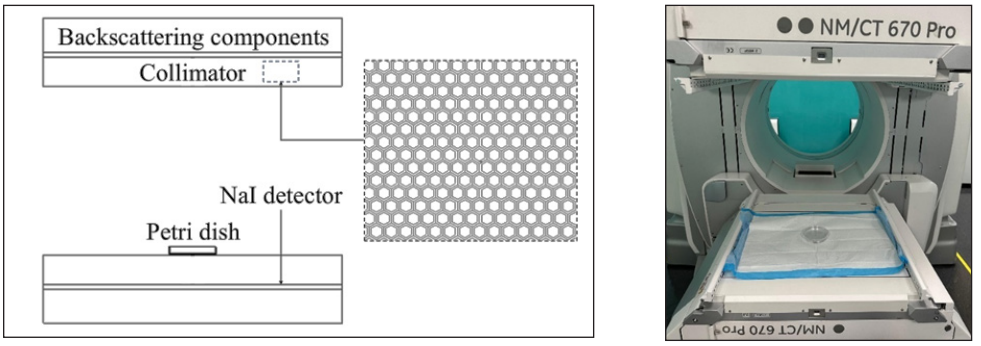
Source: Asmi et al. (2020)

System Sensitivity

Sensitivity refers to the probability that a photon will be detected by the imaging system, directly reflecting its efficiency in capturing incident radiation. To evaluate system sensitivity, planar imaging was performed using both simulation and experimental setups, where Petri dishes (d = 9 cm) were placed directly on top of the detector head, as depicted in Figure 1. Each setup contained ^{177}Lu activity concentration of 5.36 ± 0.17 MBq/ml. For the simulation configuration in MCNP5, 2.68 × 10^8 photons were simulated to replicate the same activity scanned, calculated using Equation 1:

$$n_{photons} = C \times v \times t \times (P_{113} + P_{208}) \quad [1]$$

where C is the activity concentration (5.36 ± 0.17 MBq/ml), v is the volume of the radioactive solution (20 ml), t is the acquisition time (15 s), and P_{113} and P_{208} are the emission probabilities of photons per decay for 113 and 208 keV, respectively. The experimental setup acquired planar images using a 128 × 128 image matrix with a pixel resolution of 4.42 × 4.42 mm^2.



(a) (b)

Figure 1. Planar imaging configuration of the Petri dish: (a) simulation conducted using the MCNP5 code, and (b) experimental imaging performed using the GE Discovery NM/CT 670 Pro system
Note. NaI = Sodium iodide; MCNP5 = Monte Carlo N-particle Transport, version 5; NM/CT = Nuclear medicine/computed tomography

The first component, the collimator, consists of a slab featuring hexagonal holes and lead-filled septa (Figure 1a). The next component is a detector containing sodium iodide (NaI) with a 3.67 g/cm^3 density. The following are the backscattering components, modeled using a Pyrex block measuring 6.6 cm with a density of 66% of the standard Pyrex value of 1.47 g/cm^3 (De Vries et al., 1990). This setup simulates backscattering effects from the components located behind the detector, namely, light pipe, photomultiplier tubes (PMTs), mu-metal magnetic shielding, or other structural components in an actual camera (De Vries

et al., 1990), with a particular emphasis on the impact of PMT. The surfaces and cells of the MC model were defined using the repeated structure card in MCNP5. The parameters of the MC-modeled gamma camera, as specified in the GE data sheet, are provided in Table 2. To build the MC model as realistically as possible, MCNP5 and MATLAB software were combined to achieve comparable performance to the GE HealthCare Discovery NM/CT 670 Pro.

Table 2
GE HealthCare Discovery NM/CT 670 Pro specifications (GE HealthCare, n.d.)

System parameters	Data
Collimator	MEGP
Type of collimator	Parallel hole collimator
Hole shape	Hexagonal
Hole diameter	3.0 mm
Septal thickness	1.05 mm
Hole length	58 mm
Collimator field-of-view	54 cm × 40 cm
Thickness of the crystal	9.5 mm
Total number of PMTs	59 [53 (76 mm) and 6 (38 mm)]
PMT array	Hexagonal

Note. NM/CT = Nuclear medicine/computed tomography; MEGP = Medium energy general purpose; PMT = Photomultiplier tube

As indicated in Equation 2, the decay-corrected count rate, which was derived from the sum of counts over the entire image for the corresponding acquisition time per activity as recorded in the dose calibrator, was used to compute the experimental system sensitivity (Ramonaheng et al., 2020).

$$System\ sensitivity = \frac{Count\ rate\ (cps)}{Activity\ (MBq)} \quad [2]$$

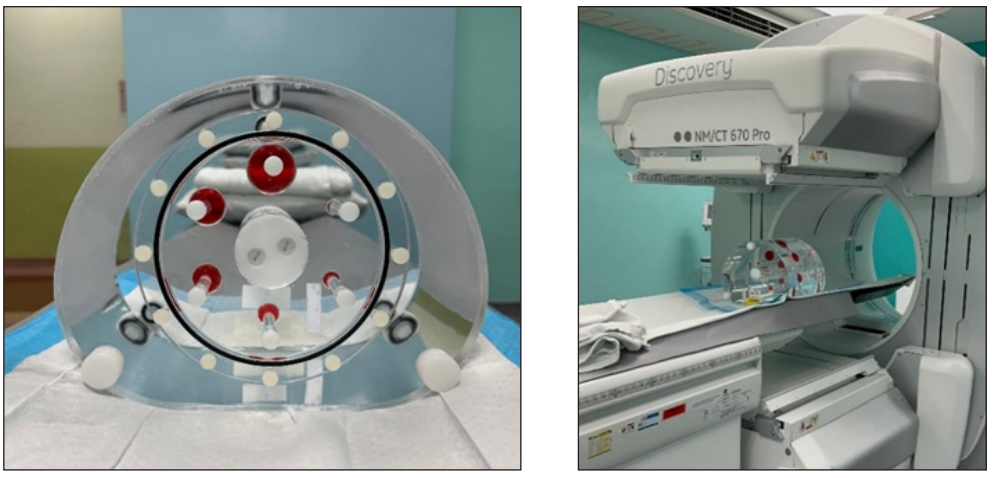
Following the approach of Morthy et al. (2025), a similar calculation was performed to determine the simulated system sensitivity. The proportionality constant, k (Bq/counts), was adopted to convert pixel values into counts, as indicated in Equation 3. Here, A represents the activity measured in Bq, while CTS denotes the total counts extracted from the energy spectrum. The pixel-wise count values in the image were then calculated by multiplying the pixel values by this constant. This method assumes a direct proportionality between activity and counts within the system's linear range. Additionally, a region of interest was defined at the center of the image to exclude edge effects (D'Arienzo et al., 2016).

$$k = \frac{A}{\sum CTS} \quad [3]$$

A comprehensive approach to model validation was provided by comparing experimental and simulated data using Pearson's correlation and Lin's concordance correlation coefficient (CCC) to assess the accuracy of the MC-modeled gamma camera.

SPECT Imaging Using Discovery NM/CT 670 Pro

A static positron emission tomography (PET) NEMA phantom (NEMA 2007/IEC 2008, Phantom Laboratory, USA) was employed for SPECT imaging using the SPECT/CT imaging system (Discovery NM/CT 670 Pro, GE HealthCare, USA) (Figure 2). The background compartment was filled with distilled water, a vial-derived 869.5 MBq of ^{177}Lu, and 500 ml of 0.9% sodium chloride (NaCl, B. Braun Medical Industries Sdn. Bhd., Malaysia). The addition of NaCl solution ensured a homogeneous radioactive solution and prevented ^{177}Lu from building up on the phantom surfaces. The spheres were filled with 47.84 ml of ^{177}Lu. A tumor to background ratio (TBR) of 5:1 was achieved, with a resulting activity concentration of ^{177}Lu at 89.64 ± 2.83 MBq/ml in the spheres and 467.00 ± 14.76 MBq/ml in the background. To mimic the human lung, 20 ml of distilled water and styrofoam were used to fill the lung insert. The lung insert simulates lung tissue and features a density similar to that of the real lung tissue of a patient. This insert creates a



(a) (b)

Figure 2. Experimental configuration of the NEMA phantom SPECT imaging: (a) NEMA phantom filled with ^{177}Lu, and (b) acquisition geometry illustrating the alignment of the phantom with the SPECT camera
Note. NEMA = National Electrical Manufacturers Association; SPECT = Single-photon emission computed tomography

non-uniform attenuation distribution in the phantom, where attenuation and scatter properties differ markedly from those of soft tissue or bone (Mia et al., 2023). Consequently, the lung insert becomes a system for rigorously evaluating the performance and accuracy of correction algorithms under varied physiological conditions, thereby validating the robustness of image-derived quantification.

Following the SPECT acquisition, the CT scan (GE Performix Ultra CT X-ray Tube, GE HealthCare, USA) was carried out at a voltage of 120 kVp and a tube current voltage range of 10–440 mA. With the system's automated exposure control, the technology minimizes potential radiation exposure while ensuring consistent image quality throughout the body. A 20% energy window was applied to all measurements and simulations, aligned with the ^{177}Lu photopeaks at 113 keV (ranging from 101 to 124 keV) and 208 keV (ranging from 187 to 228 keV). SPECT acquisitions were conducted using a step-and-shoot approach. With a 6° angular increment and a 15-s exposure time per projection, each tomographic acquisition comprised 60 projections across 360°. The data were stored in a 128 × 128 matrix with a pixel size of 4.42 × 4.42 mm. A dose calibrator (Atomlab 500, Biodex Medical Systems, USA) has been used for the activity measurement, with each measurement repeated three times to improve the statistical accuracy of the measurement. The precision of the digital readout was 0.0001 MBq/ml. According to the manufacturer's specifications for the Atomlab 500 dose calibrator, the activity measurement accuracy is within ±3% for ^{177}Lu and similar isotopes. This value, along with an estimated volume uncertainty of about ±1% (Talukdar et al., 2019), resulted in a total measurement uncertainty of 3.16% in the reported activity concentration values.

The projected images were reconstructed using the vendor's built-in quantitative software (Q.Metrix, GE HealthCare, USA), which employs the OSEM algorithm with 2 iterations and 10 subsets. The reconstruction approach included CT-based attenuation and scatter corrections to improve quantitative accuracy. Additionally, a post-reconstruction filter, namely a Gaussian filter of 4.8 mm with FWHM, was implemented to reduce noise while preserving spatial resolution.

SPECT Imaging Using MCNP5

MCNP5, developed by Los Alamos National Laboratory (USA), was utilized in this study to model the SPECT imaging system. Figure 3 illustrates the configuration of the simulated NEMA phantom and SPECT system at different rotation angles. The transformation card in MCNP5 was used to specify each rotation angle, generating 60 projection images over 360° with a 6° angular step. The spherical inserts within the phantom, containing the activity distributions, were visualized using Visual Editor (VisEd). The phantom was positioned exactly at the center of the field-of-view (FOV) of the collimator, and simulations were carried out for ^{177}Lu under conditions identical to the experimental setup. Approximately

1.25×10^{10} and 3.35×10^{8} photons were simulated for the background and spheres, respectively, to image the phantom. The sphere simulation required an average of 75 minutes per projection, leading to a total runtime of 2,325 minutes, while the background required approximately 4.375 days per projection, accumulating to 135.625 days in total on a 3.6 GHz dual-core CPU.

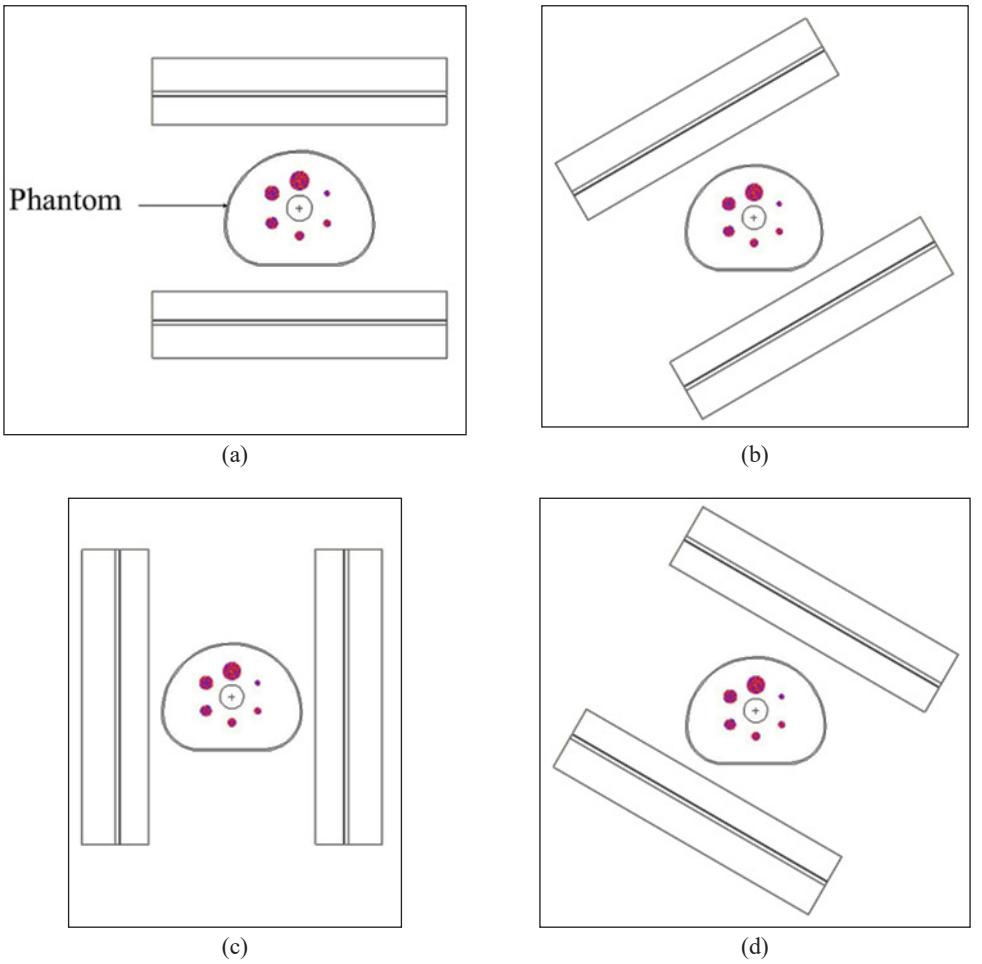


Figure 3. The geometry of the NEMA phantom and simulated SPECT system, as viewed in VisEd, is shown at different rotation angles: (a) 0°, (b) 30°, (c) 90°, and (d) 150°, with a clockwise rotation
Note. NEMA = National Electrical Manufacturers Association; SPECT = Single-photon emission computed tomography; VisEd = Visual Editor

To confine the simulation environment, a boundary was established to restrict the MCNP calculations, ensuring that any photon entering this region was excluded and forcibly terminated (Dong, 2018; Saripan, 2009). Zero importance (imp = 0) was defined for the

area beyond this boundary in MCNP5, competently removing any photon attempting to enter this area. This study acknowledges a limitation due to the unimplemented variance reduction methods.

The particle tracking (PTRAC) data generated by MCNP5 provided the core framework for the SPECT image reconstruction algorithm (Figure 4). PTRAC data, often known as list-mode data, contains essential information for photon detection, including the interaction coordinates, energy, and type of interaction. This data was created by a MATLAB algorithm to determine the energy deposited in the detector and the position of interaction. In an ideal detector, each incident photon produces a delta pulse. In practice, however, inherent variations in detector response primarily come from statistical fluctuations in the scintillation light yield. These fluctuations dominate the intrinsic resolution of the system. Although PMT gain variations contribute, their impact is secondary in scintillator-based gamma cameras. Due to these variations, the deposited energy is spread out, resulting in a Gaussian-like pulse instead of a delta pulse (Dong et al., 2018). This energy spread necessitates energy blurring in simulation to account for statistical fluctuations. A Gaussian function was applied to the energy spectrum of each detected photon, mimicking the energy spread observed in a real SPECT system to model this effect. Corresponding to the 20% window at 113 and 208 keV photopeaks, photons within the ranges of 101-124 keV and 187-228 keV were filtered. In a realistic SPECT camera, the resulting image is often blurred due to light sharing in the PMTs, causing spatial uncertainty in photon interaction positions (Dong et al., 2018). This phenomenon was modeled via position blurring, where the true interaction positions were distributed according to a 2D Gaussian distribution, leading to the formation of projections. Data necessary for image reconstruction were obtained by generating sinograms corresponding to each row of the projection image. This process is repeated for all projection angles.

The Q.Metrix application by GE HealthCare is a proprietary algorithm, precluding access to its complete internal specifications. To establish comparable reconstructions for our simulated data, an iterative reconstruction algorithm was carefully developed in MATLAB, using established quantitative SPECT reconstruction practices. The algorithm was equipped with parameters such as the number of iterations and subsets and Gaussian filter characteristics, precisely matched to those used in the experimental settings. Furthermore, attenuation and scatter corrections were incorporated into the MATLAB-based reconstruction process, comparable to the GE system. After reconstruction, the image underwent post-processing, including filtering, grayscale conversion, and transformation into uint16 format in MATLAB for improved precision before saving, ensuring compatibility with the experimental image.

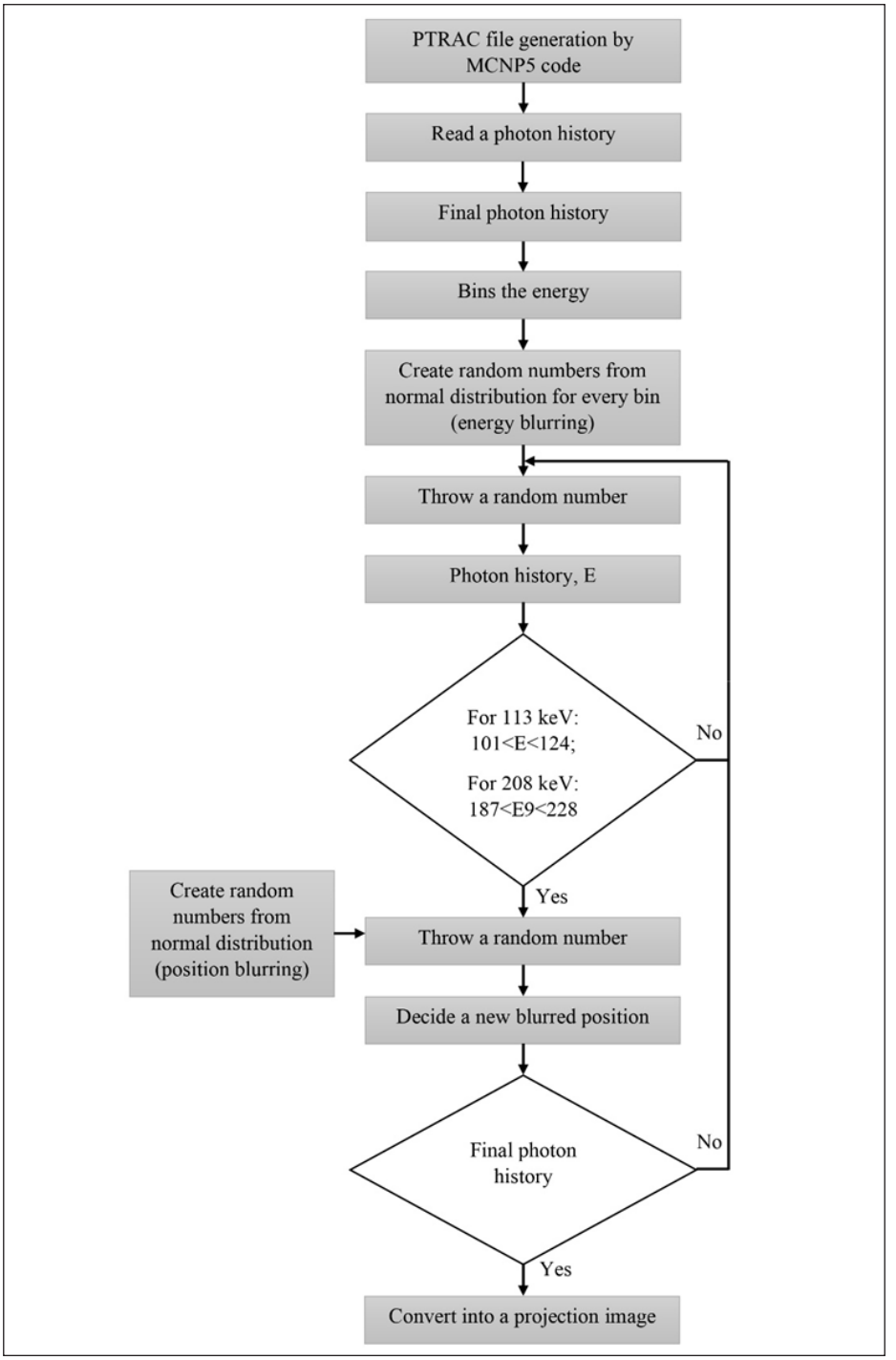


Figure 4. The SPECT image reconstruction algorithm based on PTRAC data generated by MCNP5
Note. SPECT = Single-photon emission computed tomography; PTRAC = Particle tracking; MCNP5 = Monte Carlo N-particle Transport, version 5

Verification of the MCNP5 SPECT Imaging

A set of verification criteria was used to verify the MC-simulated SPECT imaging system model through the comparison of simulated data with experimental data of the GE Discovery NM/CT 670 Pro. Verification of the MC model is necessary due to the distinct properties of the MC model. The criterion included system sensitivity, image quality, intensity profiles, and energy resolution.

Image Quality

The SSIM, which measures image quality relative to a reference image, was used in this research to evaluate the image quality. In this study, the quality of the simulated images was evaluated against the reference image, which was the experimental SPECT image. According to Roy et al. (2014) and Wang et al. (2004), SSIM is a perception-based metric that quantifies image degradation as a perceived shift in structural information. By capturing differences in contrast, brightness, and structural details, this technique assures that the evaluation is more in line with human visual perception. SSIM values vary from 0 to 1, where a value closer to 1 illustrates a high level of image similarity of simulated images in comparison with experimental images. In contrast, a value closer to 0 indicates significant structural deviation. Equation 4 illustrates the calculation of SSIM using a 3 × 3 × 3 kernel size (Rydén et al., 2021).

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad [4]$$

where x is the reference image, y is the obtained simulated image, μ_x is the average of the reference image (x), μ_y is the average of the obtained simulated image (y), σ_x is the standard deviation of the reference image, σ_y is the standard deviation of the reference image, the standard deviation of the simulated image, and σ_{xy} is the covariance of x and y. Two variables C_1 and C_2 stabilize the division with the use of a weak denominator specified as $C_1 = (K_1L)^2, C_2 = (K_2L)^2$, where L is the dynamic range of the voxel-values, and K_1 and K_2 are default values of 0.01 and 0.03, respectively. These constants ensure numerical stability, especially in areas of low contrast.

The intensity profiles across each sphere in the experimental and simulated images were analyzed to assess the spatial distribution of the reconstructed activity in the reconstructed images. These profiles were extracted through the axial slice of each sphere along the x-axis, passing through the sphere centroid. Line profile comparisons provided insights into localized intensity variations, and Pearson's correlation coefficients were used to evaluate the overall agreement between the datasets. A high correlation coefficient between the simulated and experimental intensity distributions shows that the simulation effectively

replicates the experimental imaging, verifying the reconstruction method. Additionally, FWHM values were computed from the intensity profiles using MATLAB by identifying the width at half of the maximum intensity to further characterize the differences between simulated and experimental data.

Energy Spectra and Resolution of ^{177}Lu

Energy resolution refers to a SPECT camera's ability to distinguish between photons of different energies, specifically primary and scattered radiation (Bahreyni Toossi et al., 2010; Staelens et al., 2003; Vieira et al., 2014). The energy resolutions from the simulated and experimental setups were evaluated and compared by analyzing their corresponding energy spectra. This analysis focused on the two prominent photopeaks of ^{177}Lu, 113 and 208 keV. To quantify their energy resolutions, the corresponding photopeaks were fitted with Gaussian functions using ImageJ, an open-source software given by Fiji's public domain software (Schindelin et al., 2012). The FWHM of each photopeak was derived from the Gaussian fit and measured in energy units, being the key parameter for computing the energy resolution. The energy resolution for both photopeaks was then determined using Equation 5, which ultimately demonstrates the relationship between the FWHM and peak energy. This method guarantees a uniform measurement of spectral broadening, allowing for a comparison of the performance of the gamma camera under both simulation and experimental conditions.

$$Energy\ resolution = \frac{FWHM\ (keV)}{Energy\ (keV)} \times 100\% \quad [5]$$

RESULTS AND DISCUSSION

System Sensitivity

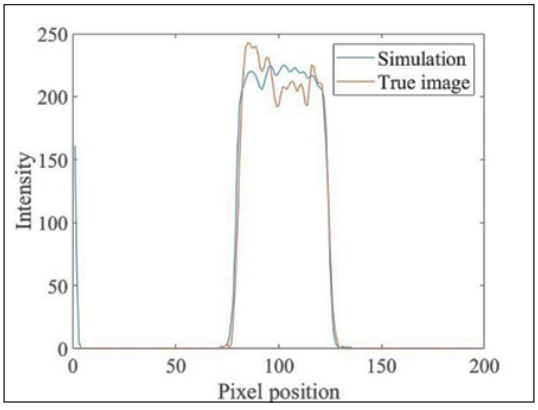
In this study, the system sensitivity determined experimentally was 12.12 cps/MBq, whereas the MCNP5 simulation yielded a value of 12.52 cps/MBq, reflecting a strong agreement with a deviation of approximately 3.34%. This small discrepancy, well within acceptable limits, underscores the reliability of the MC-modeled gamma camera for estimating activity concentration in ^{177}Lu SPECT imaging and demonstrates the robustness of the model for quantitative imaging. This verification is crucial for precise activity quantification in ^{177}Lu therapy monitoring, which is used for dosimetry calculations and treatment planning.

Figure 5 presents a visual and quantitative comparison of the simulated and experimental planar images of the ^{177}Lu-filled Petri dish. The intensity distributions of the simulated and experimental images showed a strong positive correlation, where Pearson's correlation coefficient demonstrates r = 0.9786. This implies that the simulated image closely resembles the intensity trends in the experimental data, demonstrating the MC model's

capacity to mimic spatial distribution patterns accurately. Lin's CCC was calculated, resulting in a value of 0.9780. According to McBride (2005), a CCC value greater than 0.95 indicates significant agreement, reinforcing that the MC model is highly accurate and precise in replicating the real planar imaging of ^{177}Lu. The simulation is a dependable tool for dosimetric and imaging studies since the near-unity Lin's CCC further verifies that it accurately and precisely preserves the absolute intensity values in addition to capturing variations in relative intensity.



(a) (b)



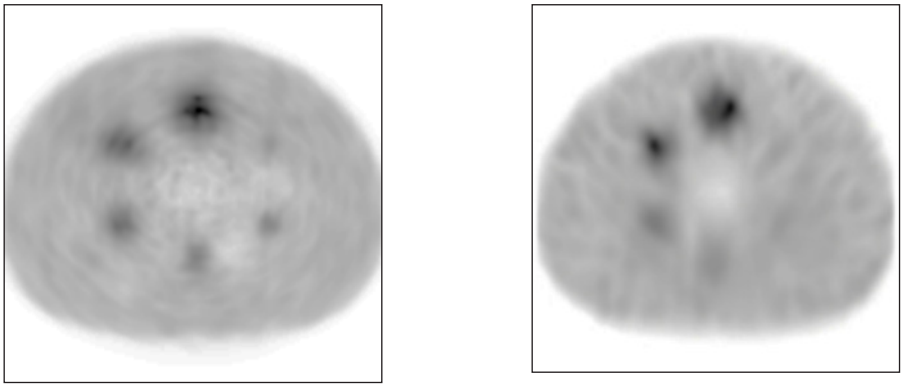
(c)

Figure 5. Comparison of Petri dish images obtained via (a) simulation in MCNP5, (b) experiment with SPECT scanner, and (c) intensity profile plotted across the center of (a) and (b)
Note. MCNP5 = Monte Carlo N-particle Transport, version 5; SPECT = Single-photon emission computed tomography

Image Quality

Figure 6 presents images reconstructed from both the MCNP5 simulation and the SPECT scanner, enabling a clear visual and quantitative comparison of the two modalities. The

SSIM was calculated to evaluate the simulated reconstruction's fidelity in comparison to the experimental data. The resulting SSIM value of 0.9348 indicates a high degree of overall structural similarity between the simulated and experimental images. This suggests that the MC model generally captures the broader structural integrity and contrast.



(a) (b)

Figure 6. Comparison of images of NEMA phantom determined via (a) simulation in MCNP5 and (b) experiment with SPECT scanner
Note. NEMA = National Electrical Manufacturers Association; MCNP5 = Monte Carlo N-particle Transport, version 5; SPECT = Single-photon emission computed tomography

However, a closer visual inspection of Figures 6a and 6b reveals some notable discrepancies between the experimental and simulated images that need further discussion, despite the high SSIM value indicating strong overall structural similarity. A distinct ring artifact appears in the simulated image but is not visible in the corresponding experimental image. This suggests that the simulation may not fully capture the complexities of scatter simulation or incorporate an ideal detector response that does not account for experimental system non-uniformities.

Furthermore, the smallest sphere (10 mm), which was not visible in the experimental image due to factors like noise and partial volume effects (PVEs), is clearly visible in the simulated image. This increased visibility of the smallest sphere implies that the model might not fully address the complex interactions of noise, PVEs, and limitations in spatial resolution present in the experimental setup. As a result, the simulation may overestimate its ability to precisely predict the detectability limits of small lesions or provide accurate quantitative results for these volumes. This could lead to inflated RCs for very small volumes.

Although the high SSIM value indicates an overall structural similarity, it is crucial to understand that this metric may not capture the clinical significance of specific localized artifacts or accurately reflect the system's true detectability limits, as shown by the

visual findings. Future refinements of the MCNP5 model will focus on addressing these specific discrepancies by refining noise models to better reflect experimental systems or implementing more advanced scatter correction algorithms within the simulation environment. This iterative refinement is vital for improving the model's fidelity and enhancing its applicability for accurate and precise quantitative analysis and patient-specific dosimetry.

Figure 7 presents a comparative analysis of intensity profiles for each sphere in the experimental and simulated images. Pearson's correlation coefficients for all spheres were computed to objectively evaluate the agreement between the two datasets. The coefficients for the spheres of sizes 13, 17, 22, 28, and 37 mm were 0.9037, 0.9515, 0.9592, 0.9626, and 0.9882, respectively. As the size of the sphere increases, these results display a progressively stronger positive correlation, suggesting that the simulated and experimental intensity distributions agree better for larger volumes. The largest sphere (37 mm) showed the highest correlation (0.9882), indicating that the simulation reproduces the activity distribution in larger spheres with minimal variation from experimental results.

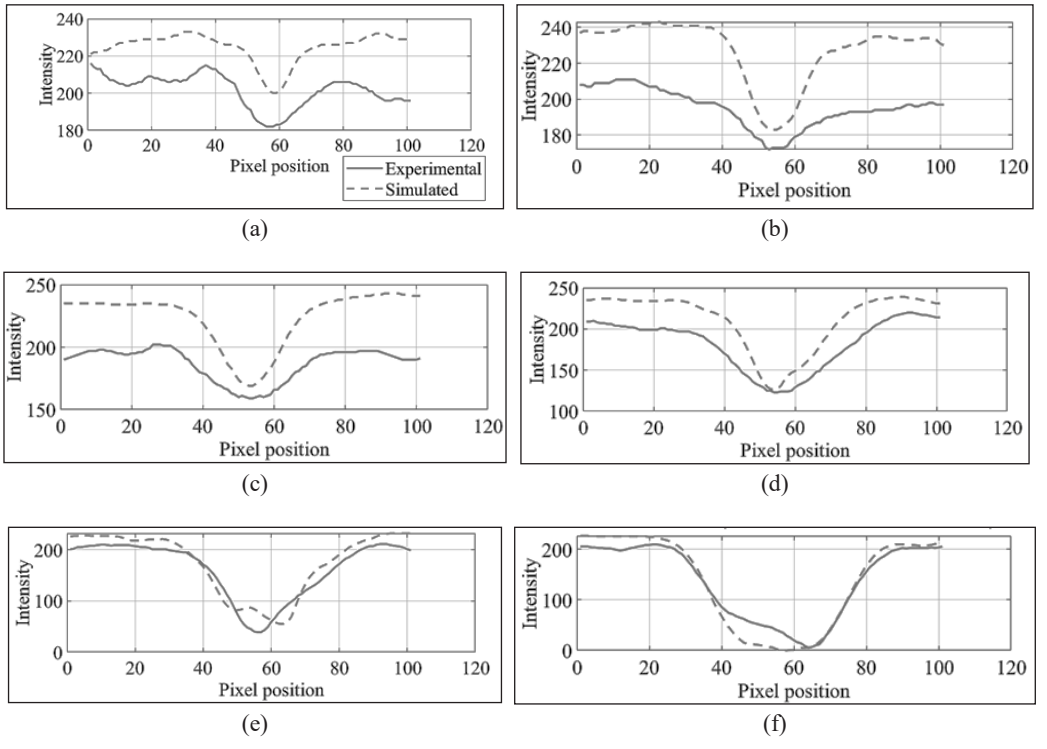


Figure 7. Comparison of intensity profiles across each sphere: (a) 10 mm, (b) 13 mm, (c) 17 mm, (d) 22 mm, (e) 28 mm, and (f) 37 mm, in experimental and simulated NEMA phantom images
Note. NEMA = National Electrical Manufacturers Association

The smallest sphere (10 mm) has a lower correlation coefficient of 0.7509. This reduced correlation highlights challenges in modeling small volume quantification. PVEs, arising from the finite spatial resolution of the gamma camera, become more pronounced as the sphere size decreases. This results in an underestimation of activity in the true volume (and thus the estimated absorbed dose) and an inaccurate volume estimation (Di Domenico et al., 2023). Although PVE affects both experimental and simulated data to a similar extent, the simulation seems to underestimate its impact. This observed discrepancy between our simulation and experiment likely stems from an overly optimistic spatial resolution model in the simulation and insufficient noise modeling. For instance, the simulation might be underestimating the intrinsic resolution degradation of the detector, or the noise characteristics of the simulated images may not fully reflect the complexities of experimental noise, thereby leading to a less pronounced PVE than observed experimentally, and artificially enhanced detectability in the simulated images.

Additionally, statistical noise, inherent to low photon counts, can significantly impact image quantification, particularly for smaller spheres (Allangba et al., 2023). Lower photon counts result in higher variability in activity reconstructions, increasing the differences between simulated and experimental images. The differences are especially notable in small volumes, where a decreased signal-to-noise ratio (SNR) complicates precise activity quantification. MC simulations often rely on idealized photon transport models that lack the exact noise characteristics of real-world imaging systems. Therefore, experimental data may show higher variations due to the statistical noise and instrumentation-related issues. This discovery is consistent with prior research, which has demonstrated that larger structures have less blurring and are less vulnerable to photon count limitations, resulting in improved quantitative accuracy (Ljungberg & Sjögreen-Gleisner, 2011; Sarrut et al., 2021; Zaidi, 1999).

Table 3 provides a quantitative summary of the FWHM values computed for each sphere in experimental and simulated data. Notably, the larger spheres (37 and 28 mm) showed excellent agreement between the experimental and simulated FWHM values, with percent differences of only 1.37 and 0.25%, respectively. For these larger spheres, the experimental and simulated intensity profiles also demonstrated good qualitative agreement in terms of overall dip shape and approximate depth.

In contrast, a pronounced discrepancy was observed for the smaller spheres (22, 17, and 13 mm). The simulated FWHM values for these spheres were consistently and significantly lower than the experimental FWHM values, implying a more idealized resolution in the simulation. Evidently, the simulated intensity profiles consistently exhibited deeper and sharper intensity dips compared to the progressively shallower and broader experimental profiles. This FWHM discrepancy was most pronounced for the 13 mm sphere, with a percent difference of 39.06%.

Table 3
FWHM values for spheres in experimental and simulated NEMA phantom images

Sphere diameter (mm)	Experimental FWHM (mm)	Simulated FWHM (mm)	Percent difference (%)
10.00	15.19	16.39	7.94
13.00	27.97	17.05	39.06
17.00	32.06	20.12	37.26
22.00	28.96	22.21	23.29
28.00	27.50	27.57	0.25
37.00	36.46	36.96	1.37

Note. FWHM = Full width at half maximum; NEMA = National Electrical Manufacturers Association

This trend reflects the nonlinear recovery behaviour in SPECT imaging and the pronounced impact of PVEs on medium-sized spheres. A previous study has shown that medium-sized spheres (13, 17, and 22 mm) are most affected by PVEs, with correction methods improving quantification by approximately 70%, while larger spheres (28 and 37 mm) exhibit minimal underestimation (Allangba et al., 2023). For the smallest sphere (10 mm), the interpretation of FWHM is complex due to its proximity to the system's intrinsic resolution limit, leading to increased noise sensitivity and fitting variability in the FWHM estimation. Furthermore, while simulated data generally show isotropic blurring (similar FWHM in x and y), experimental results reveal anisotropic effects, suggesting real-world asymmetries in the imaging system. A recent study also reported that filtering post-reconstruction images can increase the degree of negative bias in smaller objects while removing positive noise-induced bias in larger spheres of unfiltered images (Dickson et al., 2022).

Energy Resolution of ^{177}Lu

Figure 8 shows the energy spectrum from the SPECT simulation model and SPECT imaging, reflecting the energies of photons detected by the gamma camera, which are impacted by the physical properties of the radionuclide and detector. Prior to energy and position blurring, the energy spectrum exhibits discrete peaks corresponding to the energies of the emitted photons from ^{177}Lu, peaks at 113 keV and 208 keV (Figure 8a).

Gaussian energy blurring was implemented to account for system imperfections and random fluctuations in photon detection, thereby imitating real SPECT imaging (Musarudin et al., 2015). A sigma value of 1.6 was used for this Gaussian blurring applied during the iterative reconstruction process. This resulted in a smear, broadening, and Gaussian-shaped spectrum of the photopeaks (Figure 8b). While this blurring method accurately replicates the general broadening and shape of the peaks, the resulting spectrum qualitatively corresponds to the experimental energy spectrum (Figure 8c).

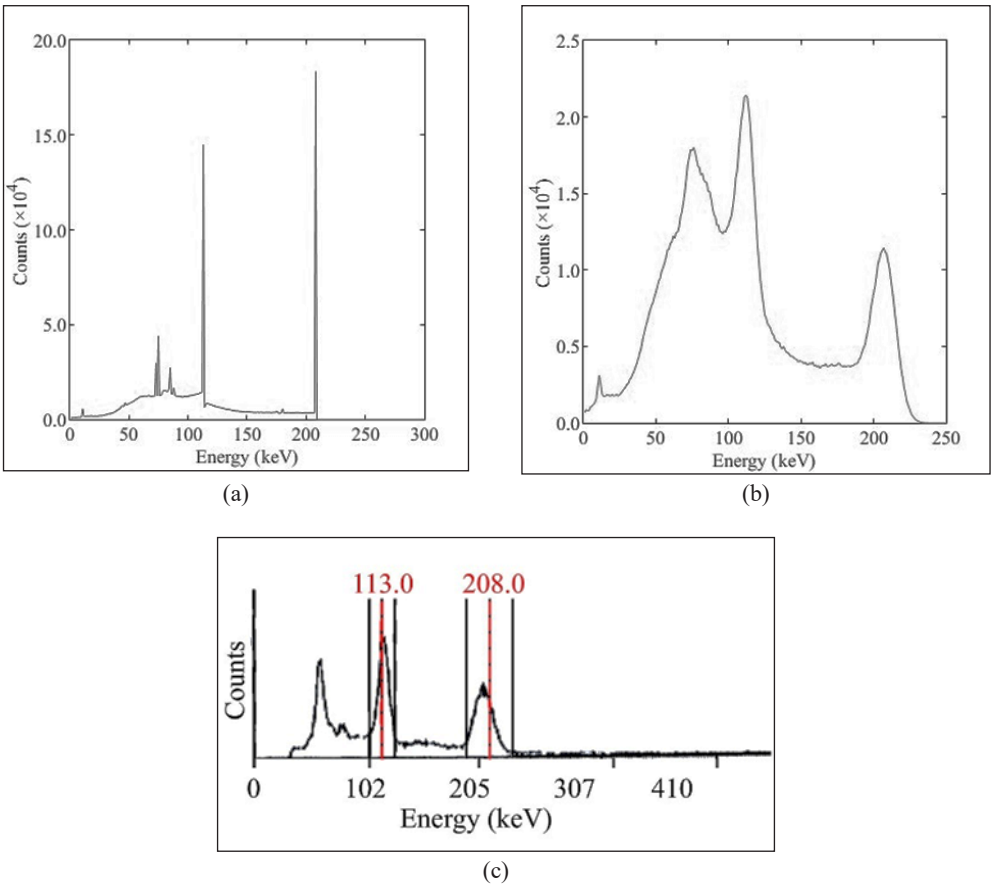


Figure 8. Simulated and experimental energy spectra: (a) and (b) represent the simulated energy spectrum before and after applying Gaussian energy blurring, respectively, and (c) presents the experimental energy spectrum
Note. Red straight lines indicate the peak energies of lutetium-177 (^{177}Lu)

The FWHM of the photopeaks was utilized to compute simulated energy resolution. The simulated energy resolution at 113 and 208 keV was 16.28 and 15.69%, respectively. This difference arises due to the energy-dependent nature of SPECT imaging detection. Lower energies, at 113 keV, typically exhibit poorer resolution compared to higher energies like 208 keV, primarily due to greater statistical fluctuations in events within the imaging chain at lower energies, where fewer photons are produced in the scintillator, resulting in broader peaks (White, 2015). The above-mentioned explains the variation in the number of simulated counts for the 113 and 208 photopeaks.

As shown in Figure 8c, the overall shape and major peaks of the simulated energy spectrum are consistent with the experimental range. However, the 208 keV peak aligns better with the experimental data compared to the 113 keV peak, where the simulated width appears wider than in the experimental data. This difference suggests that, while

the Gaussian blurring adequately models the broadening, it may not effectively represent the detector's energy response across the energy spectrum for isotopes emitting multiple energies. Previous studies have reported similar outcomes, highlighting that energy blurring does not always achieve full agreement across the entire energy spectrum for isotopes with multiple emissions (Assié et al., 2005; Morphis et al., 2021; Robinson et al., 2016). Future research on the MCNP5 model will attempt to optimize the energy blurring parameters to achieve a more precise match for all photopeaks, particularly for lower energy outputs.

CONCLUSION

This study presents a validated MCNP5 model of the GE HealthCare Discovery NM/CT 670 Pro SPECT gamma camera for ^{177}Lu SPECT imaging. The MCNP5 modeled gamma camera was systematically validated for ^{177}Lu imaging by assessing several key performance metrics, including system sensitivity, image quality, intensity profiles, and energy resolution. The results demonstrated strong agreement between the simulated and experimental system sensitivities (3.34% variation), a high SSIM value (0.9348) indicating high overall structural resemblance, and excellent correlation coefficients across larger structures (0.9882 for the 37 mm sphere and 0.9626 for the 28 mm sphere). This study thus confirms that the MCNP5 model of the SPECT scanner, along with all subsequent post-simulation programs, effectively replicates the conditions of real-world SPECT scanning for general image characteristics and larger structures.

However, the MC model revealed limitations when simulating smaller structures, particularly those under 28 mm in diameter, including underestimated blurring and overestimated contrast. The detectability of the smallest sphere (10 mm) with a lower correlation coefficient (0.7509) further highlights the need for improvement in the modeling of underlying physical interactions.

Despite these discrepancies, it is evident that MCNP5 can successfully emulate the clinical Discovery NM/CT 670 Pro, making it a reliable tool for modeling planar and SPECT imaging for ^{177}Lu, particularly for studies involving larger structures and general image characteristics. This successful validation enhances confidence in MCNP5's potential for future research. Continued refinement, focusing on optimizing noise models, detector response, and scatter simulations, is crucial to enhance the model's fidelity further. It emphasizes its greater clinical applicability, which could enhance diagnostic accuracy and activity quantification in nuclear medicine.

ACKNOWLEDGMENTS

The authors extend their sincere gratitude to the Ministry of Higher Education Malaysia for supporting this research through the Fundamental Research Grant Scheme with project code FRGS/1/2021/SKK06/USM/03/1.

REFERENCES

Allangba, K. N. P. G., Koutouan, A. K., Giuliano, A., Traoré, Z., & Traino, A. (2023). Partial Volume Effect (PVE) correction in Single Photon Emission Computed Tomography (SPECT) Imaging. *Radiation Science and Technology*, *9*(3), 26–35. https://doi.org/10.11648/j.rst.20230903.11

Asmi, H., Bentayeb, F., Bouzekraoui, Y., Bonutti, F., & Douama, S. (2020). Energy window and collimator optimization in lutetium-177 single-photon emission computed tomography imaging using Monte Carlo simulation. *Indian Journal of Nuclear Medicine*, *35*(1), 36–39. https://doi.org/10.4103/ijnm.IJNM_121_19

Assié, K., Gardin, I., Véra, P., & Buvat, I. (2005). Validation of the Monte Carlo simulator GATE for indium-111 imaging. *Physics in Medicine and Biology*, *50*(13), 3113. https://doi.org/10.1088/0031-9155/50/13/010

Bahnamiri, S. B. (2015). Investigation of collimator influential parameter on SPECT image quality: A Monte Carlo study. *Journal of Biomedical Physics and Engineering*, *5*(1), 39–48.

Bahreyni Toossi, M. T., Islamian, J. P., Momennezhad, M., Ljungberg, M., & Naseri, S. H. (2010). SIMIND Monte Carlo simulation of a single photon emission CT. *Journal of Medical Physics*, *35*(1), 42–47. https://doi.org/10.4103/0971-6203.55967

Bodei, L., Cremonesi, M., Grana, C. M., Fazio, N., Iodice, S., Baio, S. M., Bartolomei, M., Lombardo, D., Ferrari, M. E., Sansovini, M., Chinol, M., & Paganelli, G. (2011). Peptide receptor radionuclide therapy with ^{177}Lu-DOTATATE: The IEO phase I-II study. *European Journal of Nuclear Medicine and Molecular Imaging*, *38*, 2125–2135. https://doi.org/10.1007/s00259-011-1902-1

Bouchareb, Y., AlSaadi, A., Zabah, J., Jain, A., Al-Jabri, A., Phiri, P., Shi, J. Q., Delanerolle, G., & Sirasanagandla, S. R. (2024). Technological advances in SPECT and SPECT/CT imaging. *Diagnostics*, *14*(13), 1431. https://doi.org/10.3390/diagnostics14131431

Bui, M. H. H. H., Robert, A., Badel, J. N., Kvassheim, M., Stokke, C., Rit, S., & Etxebeste, A. (2023). Validation of a model of the Symbia Intevo Bold SPECT scanner for Monte Carlo simulations. In IEEE *Nuclear Science Symposium and Medical Imaging Conference* (p. 1). IEEE. https://doi.org/10.1109/NSSMICRTSD49126.2023.10338340

D'Arienzo, M., Cazzato, M., Cozzella, M. L., Cox, M., D'Andrea, M., Fazio, A., Fenwick, A., Iaccarino, G., Johansson, L., Strigari, L., Ungania, S., & De Felice, P. (2016). Gamma camera calibration and validation for quantitative SPECT imaging with ^{177}Lu. *Applied Radiation and Isotopes*, *112*, 156–164. https://doi.org/10.1016/j.apradiso.2016.03.007

De Vries, D. J., Moore, S. C., Zimmerman, R. E., Mueller, S. P., Friedland, B., & Lanza, R. C. (1990). Development and validation of a Monte Carlo simulation of photon transport in an Anger camera. *IEEE Transactions on Medical Imaging*, *9*(4), 430–438. https://doi.org/10.1109/42.61758

Di Domenico, G., Di Biaso, S., Longo, L., Turra, A., Tonini, E., Longo, M. C., Uccelli, L., & Bartolomei, M. (2023). Validation of ^{99m}Tc and ^{177}Lu quantification parameters for a Monte Carlo modelled gamma camera. *EJNMMI Physics*, *10*, 27. https://doi.org/10.1186/s40658-023-00547-6

Dickson, J. C., Armstrong, I. S., Gabiña, P. M., Denis-Bacelar, A. M., Krizsan, A. K., Gear, J. M., Van den Wyngaert, T., de Geus-Oei, L.-F., & Herrmann, K. (2022). EANM practice guideline for quantitative SPECT-CT. *European Journal of Nuclear Medicine and Molecular Imaging*, *50*, 980–995. https://doi.org/10.1007/s00259-022-06028-9

Dong, X., Saripan, M. I., Mahmud, R., Mashohor, S., & Wang, A. (2018). Characterization of SIEMENS Symbia T SPECT camera in Monte Carlo simulation environment. *Pakistan Journal of Nuclear Medicine*, *8*(1), 18–26. https://doi.org/10.24911/pjnmed.175-1540569779

Ejeh, J. E., van Staden, J. A., & du Raan, H. (2019). Validation of SIMIND Monte Carlo simulation software for modelling a Siemens Symbia T SPECT scintillation camera. In L. Lhotska, L. Sukupova, I. Lacković, & G. S. Ibbott (Eds.), *World Congress on Medical Physics and Biomedical Engineering 2018* (pp. 573-576). Springer. https://doi.org/10.1007/978-981-10-9035-6_106

Ezziddin, S., Khalaf, F., Vanezi, M., Haslerud, T., Mayer, K., Al Zreiqat, A., Willinek, W., Biersack, H.-J., & Sabet, A. (2014). Outcome of peptide receptor radionuclide therapy with ^{177}Lu-octreotate in advanced grade 1/2 pancreatic neuroendocrine tumours. *European Journal of Nuclear Medicine and Molecular Imaging*, *41*, 925–933. https://doi.org/10.1007/s00259-013-2677-3

GE HealthCare. (n.d.). *Discovery NM/CT 670 Pro: Data sheet*. https://www.probomedical.com/wp-content/uploads/2024/09/GE-Discovery-670.pdf

Ilan, E., Sandström, M., Wassberg, C., Sundin, A., Garske-Román, U., Eriksson, B., Granberg, D., & Lubberink, M. (2015). Dose response of pancreatic neuroendocrine tumors treated with peptide receptor radionuclide therapy using ^{177}Lu-DOTATATE. *Journal of Nuclear Medicine*, *56*(2), 177–182. https://doi.org/10.2967/jnumed.114.148437

Ljungberg, M., & Sjögreen-Gleisner, K. (2011). The accuracy of absorbed dose estimates in tumours determined by quantitative SPECT: A Monte Carlo study. *Acta Oncologica*, *50*(6), 981–989. https://doi.org/10.3109/0284186x.2011.584559

McBride, G. B. (2005). *A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient*. https://www.medcalc.org/download/pdf/McBride2005.pdf

Mia, M. S., Hossain, M. N., Biman, T. A., & Kabir, M. F. F. (2023). Performance evaluation of the GE Discovery IQ PET/CT system according to the NEMA NU2-2012 standard. *Bangladesh Journal of Physics*, *30*(2), 57–67. https://doi.org/10.3329/bjphy.v30i2.68384

Mishra, A. (2024). Gamma Cameras: Exploring the technology, applications, and future prospects. *American Academic Scientific Research Journal for Engineering, Technology, and Sciences*, *98*(1), 296–308.

Morphis, M., van Staden, J. A., du Raan, H., & Ljungberg, M. (2021). Modelling of energy-dependent spectral resolution for SPECT Monte Carlo simulations using SIMIND. *Heliyon*, *7*(2), e06097. https://doi.org/10.1016/j.heliyon.2021.e06097

Morthy, P. (2025). Optimization of the acquisition and reconstruction protocol for quantitative ^{177}Lu SPECT/CT. *Australasian College of Physical Scientists and Engineers in Medicine*. https://doi.org/10.1007/s13246-024-01502-0

Morthy, P., Musarudin, M., Ab Shukor, N. S., Said, M. A., Xianling, D., & Saripan, M. I. (2025). Accuracy of ^{177}Lu activity quantification using MCNP5-modeled SPECT imaging. *Applied Radiation and Isotopes*, *220*, 111786. https://doi.org/10.1016/j.apradiso.2025.111786

Musarudin, M., Saripan, M. I., Mashohor, S., Saad, W. H. M., Hashim, S., & Nordin, A. J. (2015). Validation of a clinical PET scanner using Monte Carlo simulation code: MCNP5. In *8th EUROSIM Congress on Modelling and Simulation* (pp. 36–41). IEEE. https://doi.org/10.1109/EUROSIM.2013.16

Nakanishi, K., Fujita, N., Abe, S., Nishii, R., & Kato, K. (2024). A simple method to shorten the apparent dead time in the dosimetry of Lu-177 for targeted radionuclide therapy using a gamma camera. *Physica Medica*, *119*, 103298. https://doi.org/10.1016/j.ejmp.2024.103298

Paganelli, G., Sansovini, M., Ambrosetti, A., Severi, S., Monti, M., Scarpi, E., Donati, C., Ianniello, A., Matteucci, F., & Amadori, D. (2014). 177 Lu-Dota-octreotate radionuclide therapy of advanced gastrointestinal neuroendocrine tumors: Results from a phase II study. *European Journal of Nuclear Medicine and Molecular Imaging*, *41*, 1845–1851. https://doi.org/10.1007/s00259-014-2735-5

Pells, S., Cullen, D. M., Deidda, D., Denis-Bacelar, A. M., Fenwick, A., Ferreira, K. M., Hamilton, D., Heetun, W., Julyan, P., Needham, G., Pietras, B., Price, E., Scuffham, J., Tipping, J., & Robinson, A. P. (2023). Quantitative validation of Monte Carlo SPECT simulation: Application to a Mediso AnyScan GATE simulation. *EJNMMI Physics*, *10*, 60. https://doi.org/10.1186/s40658-023-00581-4

Ramonaheng, K., van Staden, J. A., & du Raan, H. (2020). Validation of a Monte Carlo modelled gamma camera for Lutetium-177 imaging. *Applied Radiation and Isotopes*, *163*, 109200. https://doi.org/10.1016/j.apradiso.2020.109200

Robinson, A. P., Tipping, J., Cullen, D. M., & Hamilton, D. (2016). The influence of triple energy window scatter correction on activity quantification for ^{177}Lu molecular radiotherapy. *Physics in Medicine and Biology*, *61*(14), 5107–5127. https://doi.org/10.1088/0031-9155/61/14/5107

Romer, A., Seiler, D., Marincek, N., Brunner, P., Koller, M. T., Ng, Q. K. T., Maecke, H. R., Müller-Brand, J., Rochlitz, C., Briel, M., Schindler, C., & Walter, M. A. (2014). Somatostatin-based radiopeptide therapy with [^{177}Lu-DOTA]-TOC versus [^{90}Y-DOTA]-TOC in neuroendocrine tumours. *European Journal of Nuclear Medicine and Molecular Imaging*, *41*, 214–222. https://doi.org/10.1007/s00259-013-2559-8

Roy, P., Dutta, S., Dey, N., Dey, G., Chakraborty, S., & Ray, R. (2014). Adaptive thresholding: A comparative study. In *International Conference on Control, Instrumentation, Communication and Computational Technologies* (pp. 1182–1186). IEEE. https://doi.org/10.1109/ICCICCT.2014.6993140

Rydén, T., Van Essen, M., Marin, I., Svensson, J., & Bernhardt, P. (2021). Deep-learning generation of synthetic intermediate projections improves ^{177}Lu SPECT images reconstructed with sparsely acquired projections. *Journal of Nuclear Medicine*, *62*(4), 528–535. https://doi.org/10.2967/jnumed.120.245548

Saripan, M. I., Mohd Saad, W. H., Hashim, S., Mahmud, R., Nordin, A. J., & Mahdi, M. A. (2009). Monte Carlo simulation on breast cancer detection using wire mesh collimator gamma camera. *IEEE Transactions on Nuclear Science*, *56*(3), 1321–1324. https://doi.org/10.1109/TNS.2008.2012058

Sarrut, D., Etxebeste, A., Muñoz, E., Krah, N., & Létang, J. M. (2021). Artificial intelligence for Monte Carlo simulation in medical physics. *Frontiers in Physics*, *9*, 738112. https://doi.org/10.3389/fphy.2021.738112

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, *9*, 676–682. https://doi.org/10.1038/nmeth.2019

Seo, Y., Mari, C., & Hasegawa, B. H. (2008). Technological development and advances in single-photon emission computed tomography/computed tomography. *Seminars in Nuclear Medicine*, *38*(3), 177–198. https://doi.org/10.1053/j.semnuclmed.2008.01.001

Staelens, S., Strul, D., Santin, G., Vandenberghe, S., Koole, M., Asseler, Y. D., Lemahieu, I., & Van de Walle, R. (2003). Monte Carlo simulations of a scintillation camera using GATE: Validation and application modelling. *Physics in Medicine and Biology*, *48*(18), 3021–3042. https://doi.org/10.1088/0031-9155/48/18/305

Talukdar, M., Dewhirst, H., & Paulsen, A. (2019). Assaying lutetium 177 in a dose calibrator. *Journal of Nuclear Medicine*, *60*(supplement 1), 2080.

van der Zwan, W. A., Bodei, L., Mueller-Brand, J., de Herder, W. W., Kvols, L. K., & Kwekkeboom, D. J. (2015). GEP–NETs UPDATE: Radionuclide therapy in neuroendocrine tumors. *European Journal of Endocrinology*, *172*(1), R1–R8. https://doi.org/10.1530/EJE-14-0488

Vieira, L., Vaz, T. F., Costa, D. C., & Almeida, P. (2014). Monte Carlo simulation of the basic features of the GE Millennium MG single photon emission computed tomography gamma camera. *Revista Española De Medicina Nuclear E Imagen Molecular*, *33*(1), 6–13. https://doi.org/10.1016/j.remn.2013.03.009

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612. https://doi.org/10.1109/TIP.2003.819861

White, S. L. (2015). *Basic principles of gamma camera imaging and quality control*. https://amos3.aapm.org/abstracts/pdf/99-30890-359478-110770.pdf

Zaidi, H. (1999). Relevance of accurate Monte Carlo modeling in nuclear medical imaging. *Medical Physics*, *26*(4), 574–608. https://doi.org/10.1118/1.598559

Pertanika J. Sci. & Technol. 33 (6): 2771 - 2796 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Performance Evaluation of Jatropha Polyurethane Membrane Modified with Bio-based Carbon Dots for Pb(II) Removal

Umar Adam Majid[1], Zurina Zainal Abidin[1,2]*, Nur Haninah Harun[1] and Khairul Faezah Md Yunos[3]

[1]Department of Chemical and Environmental Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[2]Institute of Tropical Forestry and Forest Products, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[3]Department of Process and Food Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

ABSTRACT

The filtration efficiency of the jatropha polyurethane (JPU) membrane, enhanced with bio-based carbon dots (CDs) derived from oil palm empty fruit bunch carboxymethylcellulose, was evaluated for real-world applications, specifically for the removal of lead (Pb(II)) ions. Using a central composite design (CCD) within response surface methodology (RSM), the parameters for JPU/CD membranes with a 0.65 wt% carbon-dot loading were investigated and optimized. Pressure of range 1.5–2.5 bar, feed concentration from 100 to 200 ppm, and solution from pH (9 to 13 were fed into the RSM. An optimum Pb(II) ion removal of 82% was predicted at conditions of pH 11, 135 ppm, and 2 bar, while the validated experimental result recorded 67% Pb(II) ion removal. A statistically non-significant term was excluded from the analysis through backward elimination to improve the model's accuracy. The quadratic regression model was significant (R^2 = 0.90) for the optimization prediction. Therefore, the results from the reduction model implied a satisfactory membrane performance, offering a better prediction for Pb(II) ions removal. The addition of CDs alters the membrane structure and properties, leading to improved water permeability and Pb(II) rejection, and positioning the bio-based JPU/CDs membrane as a promising sustainable solution.

Keywords: Carbon dots, filtration, heavy metals, *Jatropha*, membrane, vegetable oil

ARTICLE INFO

Article history:
Received: 28 March 2025
Accepted: 18 May 2025
Published: 14 October 2025

DOI: https://doi.org/10.47836/pjst.33.6.16

E-mail addresses:
umaradammajid@gmail.com (Umar Adam Majid)
zurina@upm.edu.my (Zurina Zainal Abidin)
nurhaninahharun@gmail.com (Nur Haninah Harun)
kfaezah@upm.edu.my (Khairul Faezah Md Yunos)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

Water resources are important for humans as water is a basic necessity for existence. The sudden exponential growth of the human population, along with industry advancement, affects the demand for freshwater reserves. For instance, in Malaysia, the demand for freshwater was projected to increase from 15,285 m^3/d in 2010 to 20,338 m^3/d in 2020, and the freshwater reserves per capita per day are decreasing at 5.8% per year, resulting in a projection that it to be diminished (Low et al., 2016). In addition, heavy metal contamination in water sources has become a health and environmental concern due to its persistence and toxicity. Pb(II) ions are highly toxic even at low concentrations, posing serious risks to aquatic ecosystems and human health. Sources of Pb(II) pollution include industrial discharge, mining activities, battery production, and metal plating processes (Fang et al., 2022; Wan Yaacob et al., 2009; van der Kuijp et al., 2013). Conventional methods for Pb(II) removal, such as ion exchange, chemical precipitation, and adsorption, often incur high operational costs with limited efficiency and secondary waste production. As a result, membrane-based filtration systems have gained attention as an effective water purification method due to their high selectivity and energy efficiency.

Membrane filtration systems have emerged as a key technology in water and wastewater treatment, offering efficient and environmentally friendly solutions for contaminant removal. These systems operate based on selective permeability, allowing water molecules to pass through while retaining suspended solids, pathogens, and dissolved contaminants such as heavy metals and organic compounds. Additionally, membrane processes are scalable, making them suitable for both small-scale applications and large industrial systems. One advanced configuration commonly employed is the crossflow filtration system, in which the feed water flows tangentially across the membrane surface rather than directly through it. This design significantly reduces fouling by minimizing the buildup of retained particles on the membrane, enhancing both operational efficiency and membrane lifespan (Poerio et al., 2022). The shear force generated by the crossflow also helps maintain a stable permeate flux, making it ideal for long-term continuous operation.

In addition, recent advancements in bio-based membranes have provided promising solutions for removing heavy metals, emphasizing both sustainability and cost-effectiveness. Successful fabrication of green sustainable biomaterials, such as those from cellulose, vegetable oil, and others, has heightened the interest (Idress et al., 2021; Marlina et al., 2017). Jatropha oil, a renewable and biodegradable resource, has emerged as a suitable precursor for polyurethane membrane synthesis due to its high hydroxyl content and compatibility with polymerization processes. Previous work by Harun et al. (2020) reported successful synthesis of bio-based membrane from jatropha oil with and without graphene oxide (GO) and its applications for filtration (Harun et al., 2022). However, this work needs further investigation to provide an insightful understanding of

the role of additives in tailoring the properties of the synthesis of bio-based membranes. Furthermore, optimizing membrane performance for efficient heavy metal removal remains a challenge, necessitating modifications that enhance adsorption capacity, permeability, and mechanical stability.

Nanomaterials like CDs have demonstrated excellent physicochemical properties, including high surface area, tunable functional groups, and superior adsorption capabilities (Gao, 2019). Integrating CDs as additives into polymer membranes has shown potential in improving heavy metal ion removal through enhanced surface interactions and increased hydrophilicity (Harun et al., 2020; Sgreccia et al., 2024).

In this study, JPU membrane was modified with bio-based CDs derived from the carbonization of oil palm empty fruit bunch (EFB) carboxymethylcellulose (CMC). The aim here was to assess its effectiveness in removing pollutants, particularly Pb(II) removal. This research employs a central composite design (CCD) within the RSM framework to investigate the effects of key process variables, namely feed concentration, operating pressure, and pH, on Pb(II) removal performance. The findings are significant to the growing field of bio-based membrane technology and offer a sustainable approach to removing heavy metals from aqueous solutions.

MATERIALS AND METHODS

Materials and Reagents

Crude jatropha oil (CJO) and CMC of oil palm EFBs were purchased from Bionas Sdn. Bhd. (Malaysia) and Waris Nova Cove Company (Malaysia), respectively. Glacial acetic acid was obtained from Fisher Scientific (Malaysia), while Amberlite™ IR-120, hexamethylene diisocyanate (HDI), and 30% hydrogen peroxide (H_2O_2) were supplied by Sigma-Aldrich (Malaysia). Sodium carbonate (Na_2CO_3), Lead nitrate ($Pb(NO_3)_2$), and methanol were obtained from R&M Chemicals (Malaysia). Lastly, anhydrous sodium sulfate (Na_2SO_4) was attained from PC Laboratory Reagent (Malaysia).

Synthesis of CDs

The method for producing CDs followed the process obtained from Issa, Abidin, Pudza, et al. (2020). An amount of 0.2 g of fine CMC was mixed with 30ml of polyethylene glycol (PEG, Sigma-Aldrich, Malaysia, 400 of molar mass 380-420 g/mol) (Sigma-Aldrich, Malaysia). The mixture was dissolved in 25 ml of deionized water and placed into a Teflon-lined stainless steel reactor, which was then tightly sealed. The reactor was then placed in an oven (ECOCELL®, MMM Group, Germany) and heated to 270ºC for 6 hr to facilitate hydrothermal reactions. Subsequently, the liquid was cooled to room temperature and centrifuged (Biosan LMC-3000, Fischer Scientific, Latvia) at 11,200 × *g* for 15 min.

Finally, the mixture was filtered through a 2 µm Whatman filter paper to isolate the CDs from the remaining components.

Synthesis of Epoxidized Jatropha Oil (EJO), Jatropha Oil-based Polyol (JOL), JPU/CDs Membrane

EJO was synthesized via an epoxidation reaction following a previously reported method (Harun et al., 2020). CJO and glacial acetic acid were mixed in a 1 L reaction flask, with Amberlite™ IR-120 ion-exchange resin added as a catalyst at 16 wt% relative to CJO. The reaction mixture was heated to 65°C in a water bath under continuous stirring at 900 rpm. Once the temperature stabilized, 30% H_2O_2 was gradually introduced in a dropwise manner, maintaining a molar ratio of 6:1:1.7 of glacial acetic acid, carbon double bond (C=C), and H_2O_2. The temperature subsequently increased to 75°C and was maintained for 5 hr to ensure complete epoxidation. Following the reaction, the catalyst was removed, and the mixture was quenched with excess water in three consecutive cycles at 90, 20, and 90°C, respectively, facilitating efficient acid removal. The separation process resulted in the formation of two distinct layers: an upper epoxy-rich layer and a lower aqueous phase, which was discarded at the end of each cycle. To remove residual moisture, Na_2SO_4 was added at a 1:0.15 (m/m) ratio, and the mixture was dried at 80°C for 12 hr. The dried solid desiccant was then separated by decantation, yielding the final EJO product.

The obtained EJO underwent ring-opening hydrolysis to produce JOL. In a 1 L reaction flask, water (10% of the EJO weight), methanol (5:1 ratio to water), and sulfuric acid (0.3% of the total reaction mixture) were combined and heated to 64°C under continuous stirring at 800 rpm. Once the reaction temperature was stabilized, 100 g of EJO was added, and the reaction proceeded for 30 minutes. To neutralize the reaction, Na_2CO_3 was introduced, and the mixture was allowed to cool. The resulting polyol phase was separated by discarding the precipitate. Residual water and methanol were subsequently removed through vacuum evaporation at 45°C, continuing until the final JOL product was obtained, confirmed by the absence of visible aqueous residues.

CDs were incorporated into jatropha-based polymer membranes by dispersing CDs into the JOL solution using a Branson M1800 sonicator (SonicsOnline, USA). Based on preliminary screening, only the JPU/CDs membrane with 0.65 wt% CD loadings was subjected to detailed filtration evaluation. To make JPU, HDI was introduced into the mixture in a 5:5 (v/v) ratio with JOL (Harun et al., 2020). The reaction mixture was stirred at 600 rpm while maintaining a temperature range of 90–100°C for 1 hr to ensure homogeneous mixing and polymerization. The resulting solution was cast onto a glass Petri dish, forming a membrane layer with a thickness of 0.5 mm, as measured using a caliper. The membrane was then subjected to thermal curing in an oven (ECOCELL®, MMM Group, Germany) at 150°C for 8 hr. Upon completion of the curing process, the

JPU/CDs composite membrane was carefully exfoliated from the Petri dish using a spatula and used in filtration.

Filtration Method

A crossflow system was used to perform the filtration process to test the JPU/CDs (0.65 wt%) membrane. The results from the filtration process were analyzed using inductively coupled plasma optical emission spectroscopy (ICP-OES, Optima 7300 DV PerkinElmer, USA) to measure the amount of remaining Pb(II) after the filtration. Results obtained from the ICP were then computed to determine the flux (Jw) and the amount of Pb(II) ions rejected (R), using Equations 1 and 2.

$$J_w = \frac{V}{A \cdot \Delta t} \quad \textbf{[1]}$$

where V is the permeate (L), A is the effective surface area (m^2), J_w is the water flux ($L/m^2.h$), and t is the time taken for filtration (min).

$$R = \left(1 - \frac{C_p}{C_f}\right) \times 100\% \quad [2]$$

where C_p and C_f are the Pb(II) ion concentrations in the permeate and the feed, respectively.

Screening Parametric by Single Factor Experiment Studies

In this experiment, preliminary screening was conducted to determine the appropriate range of factors that influence the experiments, which was essential to decide the best range to be entered in RSM. Three different factors, acidity, pressure, and concentration, were tested to find the optimal conditions, with the corresponding studied ranges of 100–250 ppm, 1.5–2.5 bar, and pH 9–13, respectively. The tests examined the effects of each component on the removal of Pb(II) ions, varying one factor at a time (OFAT), within their corresponding parameter ranges. Then, based on these results, the ranges were verified to be appropriate for optimization studies.

Optimization by RSM

Pb(II) ion feed concentration, solution pH, and pressure were three independent variables, used in RSM to create and optimize the experimental design for optimizing Pb(II) removal from the aqueous solution (Table 1) and the interaction between variables. From the CCD, 17 trial experiments were performed, with three repetitions at the center points. The CCD

was used to reduce the number of experiments, as it can gather information on data fitting with fewer experimental trials for the optimization process (Chee et al., 2021; Harun et al., 2022).

Table 1
RSM experimental design for Pb(II) removal using JPU/CDs 0.65 wt% membrane

Symbol	Parameters	Low	High
X_1	Concentration (ppm)	100	200
X_2	Pressure (bar)	1.5	2.5
X_3	Acidity (pH)	9	13

Note. RSM = Response surface methodology; Pb(II) = Lead ions; JPU = Jatropha polyurethane; CDs = Carbon dots; X_1 = Concentration; X_2 = Pressure; X_3 = Acidity

Then the response was fitted into a second-order polynomial model, as shown in Equation 3:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k} \sum_{j=i+1}^{k-1} \beta_{ij} x_i x_j \quad [3]$$

where y is the dependent variable (Pb(II) rejection); x_i and x_j are the independent variables that influence the response y; β_0, β_i, β_{ii}, and β_{ij} are the coefficients of intercept, linear term, quadratic term, and interaction, respectively, and k is the number of variables. In this case, $k = 2$.

Determination of Zeta Potential at Different Solution pH Values

A particle size analyzer, Zetasizer (LA-960V2, HORIBA, Japan), was used to measure the zeta potential of the sample. The solution pH was adjusted from 9 to 13 using sodium hydroxide (NaOH, 1.0 mol/L, Sigma-Aldrich, Malaysia). The zeta potential, which arises as an electrical double layer developed at the solid–liquid interface, describes the surface charge of a membrane. It is influenced by the functional groups on the membrane surface, the ions in the feed solution, and the pH (Deshmukh & Childress, 2001; Hu et al., 2023).

RESULTS AND DISCUSSION

Screening Factors by OFAT Experiments

Preliminary screening through OFAT was conducted to determine the range of parameters affecting the performance of Pb(II) removal. These ranges served as a feeder for the subsequent optimization work using RSM. The range of values studied was as follows:

(a) feed concentration (100, 150, 200, and 250 ppm); (b) water pressure (1, 1.5, 2.0, and 2.5 bar); and (c) pH values (3, 4, 5, 9, 11, and 13).

Effect of Feed Concentration on Pb(II) Rejection

In studying the effect of feed concentration, experiments were conducted at a constant pH of 7 and a pressure of 1.5 bar. Based on the results obtained, as shown in Figure 1, the rejection trend initially increased until it reached a peak of 32% at 150 ppm, and then decreased as the initial feed concentration was further increased. This may be explained by changes in membrane charge density that can be affected by the bulk ion concentration. This influences both the selective ion adsorption on the top membrane surface and within the pore. The build-up of ions near the membrane surface increases the membrane charge density and thus increases the rejection of Pb(II). However, the excess concentration of solute on the surface of the membrane can repulse the solvent from passing through the membrane. This scenario can be attributed to concentration polarization, in which the solute accumulates at the membrane surface, thereby increasing the local osmotic pressure and reducing Pb(II) rejection (Idress et al., 2021). Hence, the 100 to 200 ppm concentration range was selected for implementation in the optimization using RSM.

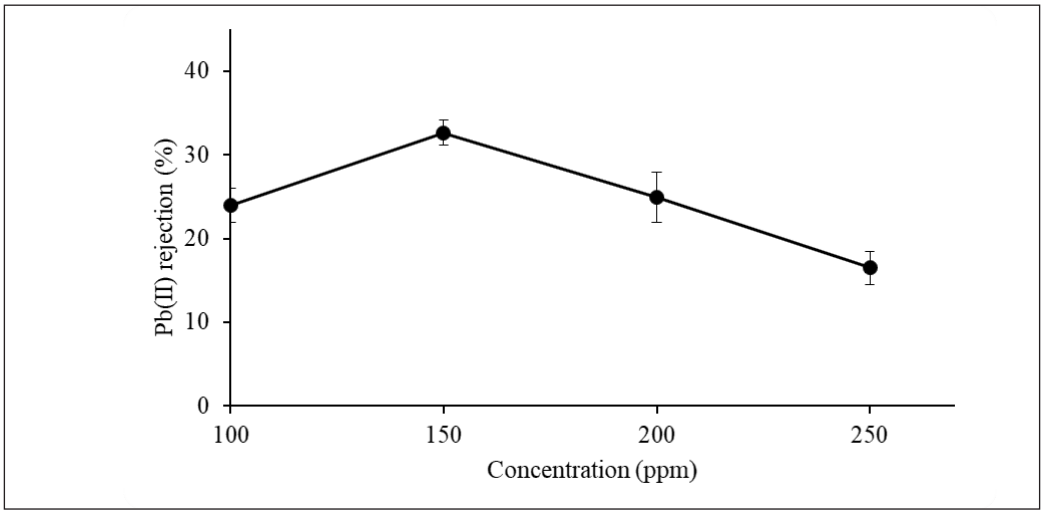


Figure 1. The lead ions (Pb(II)) rejection at different feed concentrations

Effect of Feed Pressure on Pb(II) Rejection

For studying the effect of pressure, the value was tested from 1 to 2.5 bar with an increment of 0.5 bar at a constant feed concentration of 100 ppm and pH 7. The rejection of Pb(II) increased as the pressure increased, reaching its highest value at 2.5 bar (Figure 2). The peak of the rejection was not found due to the system operation constraints and limitations.

Therefore, the pressure range of 1.5–2.5 bar was used in the following optimization process. High pressure exerts more force to carry solute to the membrane surface and causes solute accumulation on the membrane surface, which may contribute to concentration polarization (Emamjomeh et al., 2019). Meanwhile, the increased variability observed in the error bars at higher operating pressures (2.0 and 2.5 bar) can be attributed to intense mechanical stress on the membrane, potentially causing minor structural shifts or compaction, which can lead to inconsistent flow rates and separation performance (Stade et al., 2013). It may also be due to the inherent randomness and heterogeneity of the bio-based material used in membrane fabrication (Morales-Jiménez et al., 2023).

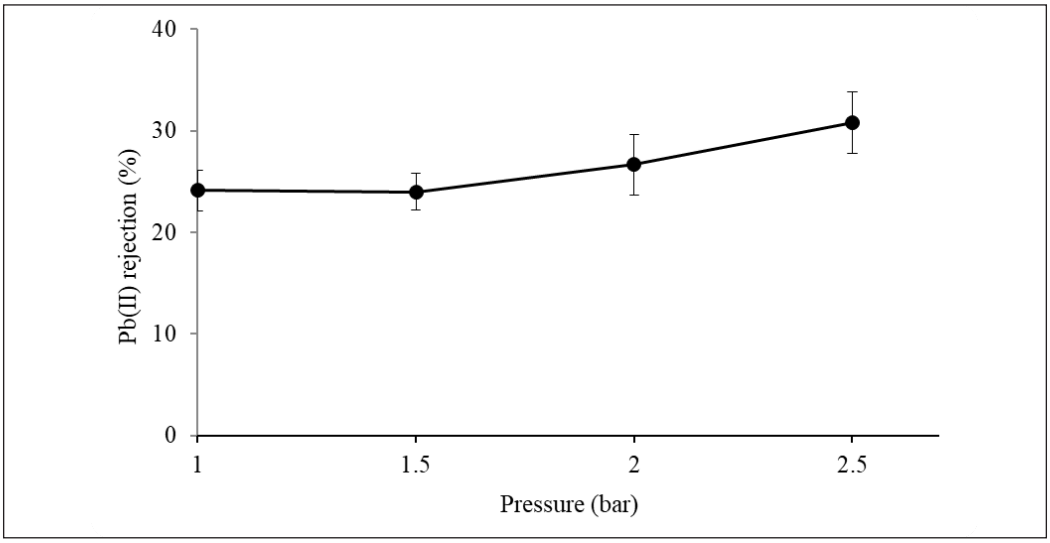
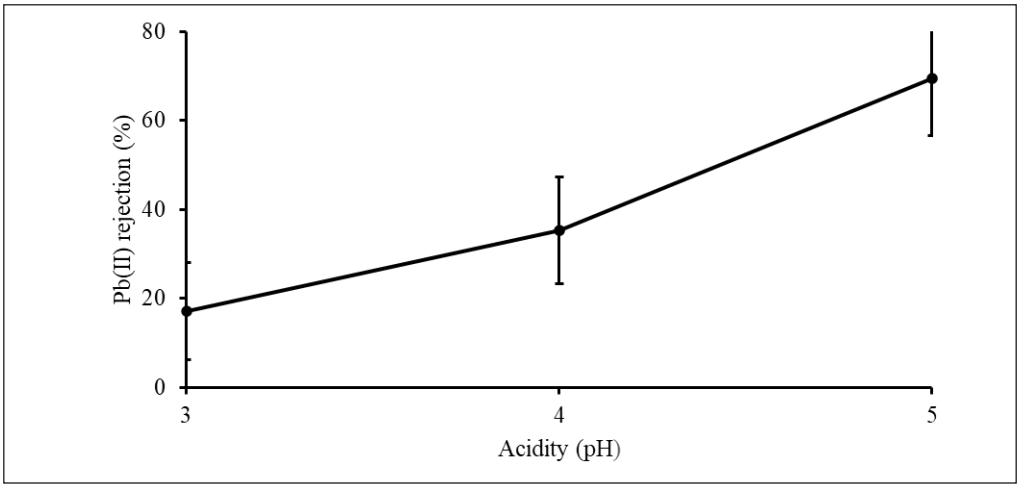


Figure 2. The lead ions (Pb(II)) rejection at different feed pressures

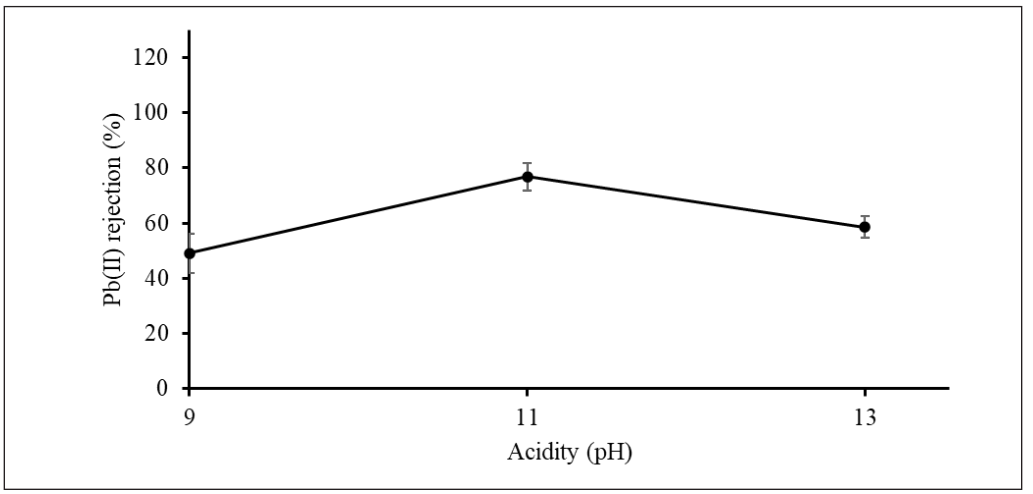
Effect of Solution pH on Pb(II) Rejection

A constant pressure of 1.5 bar and a feed concentration of 100 ppm were selected to investigate the effect of pH. The investigation in acidic conditions was conducted from pH 3 to pH 5, showing an increment from 17 to 69% rejection at pH 3 and pH 5, respectively, with no optimum peak observed (Figure 3a). Further testing in basic conditions showed an increase in the rejection from pH 9 to pH 11, and started to decrease at pH 13. The optimum peak was only found at pH 11 at 76% rejection (Figure 3b). Although the error bars at pH 4 and 5 may appear more prominent visually, the standard deviations at pH 3, 4, and 5 are relatively similar, ranging from approximately 11 to 13%. This indicates that the variability in measurements across these pH values is consistent and within an acceptable range. The apparent difference in error bar size is not due to increased variability at specific pH levels but rather sometimes reflects the normal fluctuation expected from the use of bio-based

membrane materials, which can exhibit minor structural or compositional differences across samples. pH values ranging from 6 to 8 were excluded from this study, as operating near neutrality better reflects natural waters and minimizes charge-related interferences that can compromise membrane performance. Previous work has shown that polymer membranes perform most effectively under strongly acidic or basic conditions, where surface charges are maximized (Harun et al., 2022; Khamis et al., 2023). In the near neutral range (pH 6–8), however, the zeta potential approaches the membrane's isoelectric point, lowering surface-charge density and, consequently, diminishing electrostatic interactions with Pb(II) ions (Duclos-Orsello et al., 2004).



(a)



(b)

Figure 3. The Pb(II) rejection at (a) acidic and (b) basic conditions, respectively

Because pH alters the membrane's surface-charge density and thus its zeta potential, the filtration mechanism and overall efficiency are correspondingly affected. To clarify this relationship, the zeta potential of the JPU/CDs membrane in both distilled water and an aqueous Pb(II) solution across the relevant pH range was measured. As shown in Figure 4, the JPU/CDs 0.65 wt% membrane in distilled water exhibits a negative zeta potential of -41, -32, and -32 mV at pH 9, pH 11, and pH 13, respectively. This indicates that the JPU/CDs membrane possesses a negatively charged surface under alkaline conditions. Generally, at high pH, the concentration of H^+ ions decreases, leading to the deprotonation of functional groups such as hydroxyl (-OH) and carboxyl (-COOH), thereby increasing the negative charge on the membrane surface. Conversely, under low pH conditions, the abundance of H^+ ions promotes protonation of these functional groups, resulting in a more positive surface charge. However, at pH 11 to 13, zeta potential started to plateau as the surface functional group has fully deprotonated, nearing surface saturation (Junker et al., 2023).

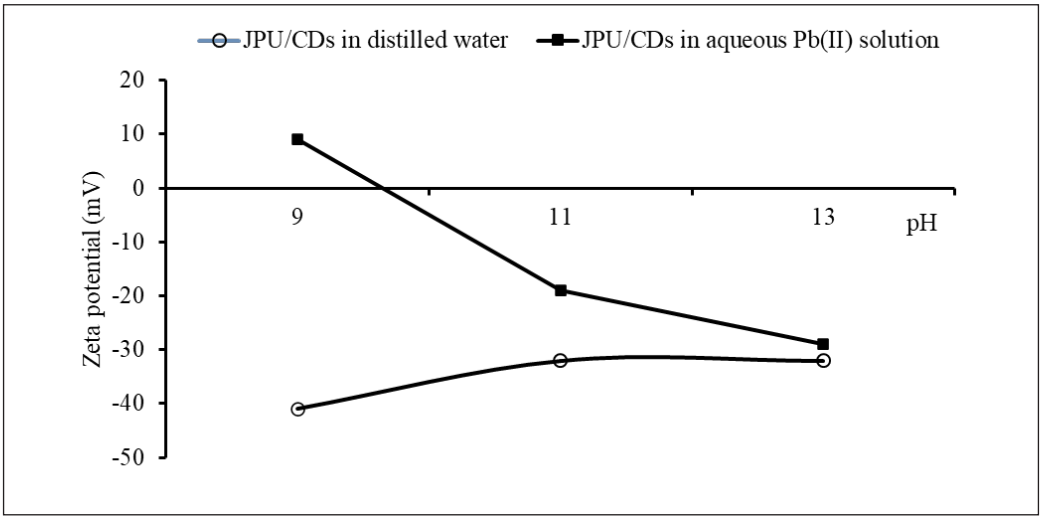


Figure 4. Zeta potential on jatropha polyurethane (JPU)/carbon dots (CDs) 0.65 wt% membrane at different pH levels in distilled water and in lead ions (Pb(II)) solution

In aqueous Pb(II) solution, the zeta potential values were measured to be 9, -19, and -29 mV at pH 9, pH 11, and pH 13, respectively. Notably, at pH 9, the value significantly shifted from -41 mV (in distilled water) to 9 mV (in Pb(II) solution). This can be attributed to the charge reversal caused by the adsorption of Pb(II) ions onto the negatively charged membrane surface. Under basic conditions, deprotonation of surface functional groups increases the negative surface charge density, facilitating strong electrostatic attraction between the membrane and lead (Pb^{2+}) ions. As Pb^{2+} ions accumulate and bind to these sites, they effectively neutralize or even overcompensate for the negative charge, leading to a net positive zeta potential. (Figures 5 and 6).

As the pH increased, with the progressive addition of NaOH, the zeta potential changed back toward more negative values (-19 mV at pH 11 and -29 mV at pH 13). This trend suggests that at higher pH, Pb(II) ions may experience precipitation as lead hydroxide ($Pb(OH)_2$) and reduced electrostatic interactions with the membrane surface, thereby restoring the negative charge (Figures 5 and 6). Additionally, the presence of excess hydroxyl (OH^-) ions may have contributed to charge repulsion effects, leading to a more negative zeta potential. However, at pH 13, the competition between Pb(II) and counter ions (such as sodium ions [Na+] and others) screens the negatively charged membrane surface, reducing the electrostatic attraction of Pb^{2+} and thereby lowering Pb(II) removal efficiency (Figure 6). This highlights the pH-dependent nature of the membrane's surface charge in the removal of heavy metals.

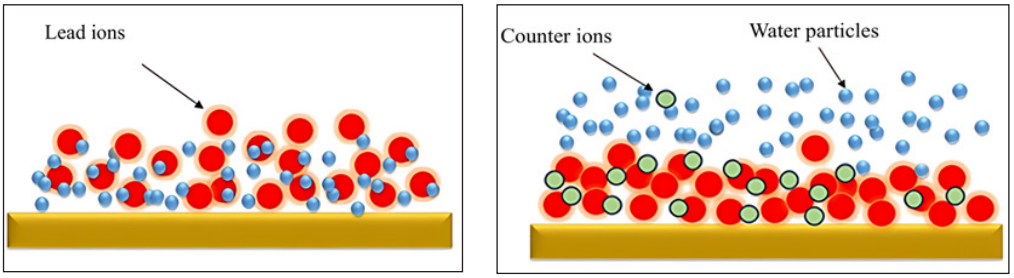


Figure 5. Illustration of the mechanisms where (a) favorable adsorption of lead ions (Pb(II)); (b) concentration polarization on the membrane surface

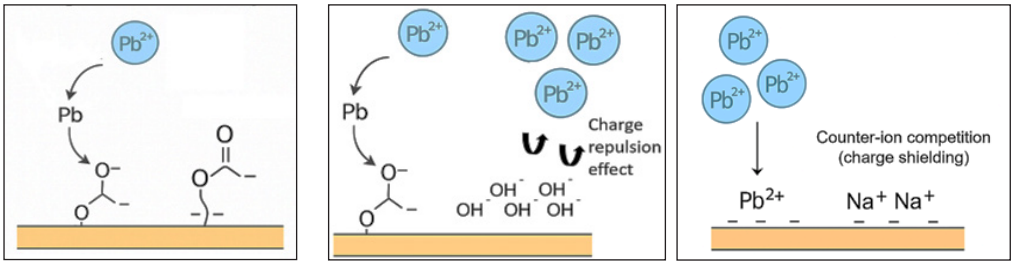


Figure 6. Schematic illustration of the possible binding mechanism at pH 9, pH 11, and pH 13, respectively
Note. Pb^{2+} = Lead ions; OH^- = Hydroxide ions; Na^+ = Sodium ions

CDs have long been recognized as effective polymer modifiers, substantially enhancing physical, chemical, and thermal stability (Feng et al., 2021). These improvements impart the material with distinctive functional attributes, which in turn promote high rejection efficiencies for target compounds (Keçili et al., 2020). In addition, the presence of abundant

hydroxyl groups on CDs increases the membrane's hydrophilicity and adsorption capability, enabling stronger interactions with lead ions via hydrogen bonding or coordination (Petrović, 2008; Shabbir et al., 2023; Xia et al., 2023). The use of additives like CDs also modifies surface characteristics and pore structures, yielding a more uniform distribution of smaller pores that enhances the size exclusion of heavy metal ions (Cruz et al., 2023; Harun et al., 2022; Khdary et al., 2023; Xia et al., 2023). Altogether, these enhancements result in a membrane with superior rejection capabilities for heavy metals, especially Pb(II) ions in this study, and improved overall operational efficiency.

Membrane fouling is an inherent phenomenon in pressure-driven separations, inevitably diminishing filtration performance. While this study did not explicitly quantify fouling, the progressive decline in permeate flux serves as indirect evidence, indicating the increasing hydraulic resistance caused by solute deposition on or within the membrane during operation, as similarly reported by other studies (Abu-Zurayk et al., 2023). This accumulation is expected to impact long-term membrane performance by reducing available binding sites and potentially increasing filtration resistance. The retained contaminants are expected to contribute to progressive fouling of the filtration medium (Al-Rashdi et al., 2013; Shi et al., 2014; Zularisam et al., 2007). On the other hand, the improved hydrophilicity imparted by CDs' functional groups has been reported to facilitate better water permeability while reducing fouling, thus supporting sustained filtration performance (Mamba et al., 2021; Xia et al., 2023).

In short, based on OFAT studies, the suitable range for performing RSM optimization analysis was effectively determined to be as follows: feed concentration (100, 150, and 200 ppm), pressure (1.5, 2.0, and 2.5 bar), and pH values (9, 11, and 13).

Optimization of JPU/CDs Membrane Filtration

The experiment was further optimized by utilizing the OFAT findings, which were then fed into the RSM software to investigate the relationship between all the parameters. RSM software implemented a mathematical model (CCD) to minimize the amount of testing needed to achieve the output. The input parameters were feed concentration (ppm) (A), pressure (bar) (B), and acidity (pH) (C). The results are shown in Table 2.

Table 2
RSM simulation data when utilizing CCD and face-centered alpha

No.	Concentration (ppm)	Pressure (bar)	Acidity (pH)	Pb(II) rejection (%)
1	100	1.5	9	16
2	200	1.5	9	45
3	150	2	9	62
4	100	2.5	9	41

Table 2 *(continue)*

No.	Concentration (ppm)	Pressure (bar)	Acidity (pH)	Pb(II) rejection (%)
5	200	2.5	9	53
6	150	1.5	11	43
7	100	2	11	56
8	150	2	11	91
9	150	2	11	80
10	150	2	11	98
11	200	2	11	63
12	150	2.5	11	64
13	100	1.5	13	51
14	200	1.5	13	19
15	150	2	13	69
16	100	2.5	13	66
17	200	2.5	13	14

Note. RSM = Response surface methodology; CCD = Central composite design; Pb(II) = Lead ions

Model Fitting

Using RSM, experimental data were analyzed by fitting them into four different models: linear, two-factor interaction (2FI), quadratic, and cubic, to determine the regression equation. The models were evaluated based on the *F*-test value and *p*-value obtained from analysis of variance (ANOVA), as summarized in Table 3. The fitting process was based on polynomial models, including linear, quadratic, and cubic equations. In general, model selection was guided by the highest-order polynomial that was not aliased (Behera et al., 2018; Issa, Abidin, Sobri, et al., 2020).

Table 3
RSM response data using CCD using face-centered alpha

Source	Sum of squares	Df	Mean square	*F*-value	*p*-value Prob > *F*	Remarks
Mean	50996.37	1	50996.37	-	-	
Linear	534.83	3	178.28	0.27	0.8479	
2Fl	2197.60	3	732.53	1.13	0.3827	
Quadratic	5594.76	3	1864.92	14.82	0.0020	Suggested
Cubic	199.86	4	49.97	0.22	0.9112	Aliased
Residual	681.18	3	227.06	-	-	
Total	60204.62	17	3541.45	-	-	

Note. RSM = Response surface methodology; CCD = Central composite design; Df = Degree of freedom; Prob = Probability; p-value < 0.05 denotes significance; The underline means quadratic is the most fitted and the selected model

Additionally, to further assess model compatibility, the lack-of-fit (LOF) value was evaluated. The reliability of the model depends on the significance of the LOF value, which is influenced by midpoint replication. A model is considered reliable when the LOF value is non-significant (p-value > 0.005). Based on this assessment, the quadratic model, with a LOF value of 0.4061, was identified as the best fit for the experimental data and deemed non-significant.

ANOVA

To evaluate the significance of the fitted model, an ANOVA was conducted. As shown in Table 4, the coefficient of determination (R^2) for the predicted model is 0.9043, indicating a strong correlation between the experimental data and the model predictions. Since R^2 is close to 1, it suggests a good model fit. However, R^2 alone may not be a reliable measure

Table 4
RSM response data using CCD using face-centered alpha

Source	Sum of squares	Df	Mean square	*F*-value	*p*-value Prob > *F*
Model	8327.197	9	925.244	7.351	0.008
A - Concentration	130.287	1	130.287	1.035	0.343
B - Pressure	403.528	1	403.528	3.206	0.116
C - pH	1.01792	1	1.018	0.008	0.931
AB	164.294	1	164.294	1.305	0.291
AC	1966.542	1	1966.542	15.624	0.005
BC	66.765	1	66.765	0.530	0.490
A^2	604.735	1	604.735	4.805	0.065
B^2	1108.574	1	1108.574	8.807	0.021
C^2	213.961	1	213.961	1.699	0.234
Residual	881.047	7	125.864		
Lack of fit	715.297	5	143.059	1.726	0.406
Pure error	165.750	2	82.875		
Cor total	9208.244	16			
Statistical analysis of the regression equation					
R-squared	0.904		Std. Dev.	11.22	
Adj *R*-squared	0.781		Mean	54.77	
Pred *R*-squared	0.579		CV%	20.48	
Adeq precision	8.189		PRESS	3870.81	

Note. RSM = Response surface methodology; CCD = Central composite design; Df = Degree of freedom; Prob = Probability; AB = Concentration-pressure; AC = Concentration-pH; BC = Pressure-pH; Adj *R*-squared = Adjusted R-squared; Pred R-squared = Predicted R-squared; Adeq precision = Adequate precision; Cor total = Corrected total; Std. Dev. = Standard deviation; CV = Coefficient of variance; PRESS = Residual error sum of squares

of model quality, as it can be artificially inflated by adding additional terms, regardless of their statistical significance.

To account for this limitation, adjusted R^2 (adj-R^2) and predicted R^2 (pred-R^2) provide more robust assessments of model performance (Halim et al., 2021). Adj-R^2 accounts for the number of predictors in the model, ensuring that only significant terms contribute meaningfully to the model. Meanwhile, predicted R^2 evaluates the model's predictive capability, helping to identify potential overfitting.

In this study, the adj-R^2 value of 0.7813 suggests that the model effectively describes the relationship between experimental data and model predictions. However, the pred-R^2 value of 0.5796 indicates a potential need for model refinement, as a lower pred-R^2 may suggest overfitting or the presence of non-significant terms that should be reconsidered. Therefore, further model optimization, such as reducing non-essential terms, may be necessary to improve its predictive accuracy.

In addition to that, the model's validity was improved based on other statistical characteristics such as adequate precision (Adeq precision), coefficient of variations (CV%), and residual error sum of squares (PRESS). The results of these data are shown above in Table 4. The Adeq precision assesses the signal-to-noise ratio that compares the range of the predicted values at the design points to the average prediction variance (Halim et al., 2021). That is, the ratio higher than 4 shows adequate model discrimination. The Adeq precision value obtained is 8.189, suggesting that the ratio is adequately accurate. In addition, the CV% value represents the degree of dispersion of the data points around the mean value. A lesser value (10%) indicates good reproducibility of the model. However, the obtained CV% from this model is 20.48. This shows the standard deviation is high relative to the mean. Apart from that, the PRESS value describes the model's capacity to fit each point in the design by fitting the model to all design points, excluding the predicted one (Halim et al., 2021). The residual values are then squared and added to yield the PRESS value reported in ANOVA. A lower value is preferred, indicating a high pred-R^2 score. However, the PRESS value generated from this is 3870.81, which is quite large, and can be associated with the low pred-R^2 value.

To further enhance the accuracy of the predicted model, additional steps were taken to eliminate a statistically non-significant term from the analysis using a reduction method. This process was carried out incrementally through backward elimination, starting with the least significant term, until a higher pred-R^2 value was achieved. The analysis revealed that removing terms with a p-value greater than 0.5 provided the best prediction for Pb(II) ion rejection. As a result of this reduction method, the pred-R^2 value improved to 0.8878 (88.78%), indicating strong predictive capability for the refined model.

Additionally, the adj-R^2 value saw a slight increase to 0.7941 (79.41%). The reduced model also showed improvements in the LOF and adequate precision, with values rising

to 0.438 and 8.532, respectively. Furthermore, the CV% and PRESS values decreased to 19.87% and 3572.71, respectively. These results demonstrate that the model's precision and predictive performance were enhanced. The final empirical models, after excluding the insignificant terms (BC) and (y), are presented in Equation 4. This equation highlights the inclusion of both linear effects and interactions between parameters for predicting Pb(II) ion rejection.

$$y = 80.83 - 3.61\ (A) + 6.35\ (B) + 0.32\ (C) - 4.53\ (AB) - 15.68\ (AC) - 15.02\ (A^2) - 20.34\ (B^2) - 8.94\ (C^2) \quad [4]$$

where *A* represents the feed concentration (ppm), *B* denotes pressure (bar), and *C* stands for the solution's pH value. A positive sign in the equation signifies a synergistic effect, whereas a negative sign reflects a resistive effect, based on the factors involved.

In the above equation, the coefficient of *A* (feed concentration) is -3.61, indicating that an increase in Pb(II) feed concentration negatively impacts removal efficiency. This can be attributed to saturation of adsorption sites on the membrane, where excessive Pb(II) ions exceed the membrane's capacity, reducing overall rejection. The coefficient of *B* (pressure) is +6.35, suggesting a positive correlation between pressure and Pb(II) removal. Increased pressure enhances convective mass transport, driving more Pb(II) ions toward the membrane and improving rejection efficiency through size exclusion and electrostatic repulsion. Finally, the coefficient of *C* (pH) is +0.32, indicating a minor positive effect of pH on Pb(II) removal. At higher pH, Pb(II) may form insoluble hydroxides ($Pb(OH)_2$), facilitating precipitation and improving removal. However, since the effect size is small, the membrane's role in rejection may dominate over precipitation.

The negative interaction between feed concentration and pressure (coefficient -4.53) implies that at higher Pb(II) concentrations, increased pressure does not proportionally enhance removal efficiency. This could be due to concentration polarization effects, where excessive Pb(II) ions accumulate near the membrane surface, creating a diffusion barrier that limits the process. The strong negative effect of the feed concentration–pH interaction (-15.68) suggests that increasing pH at higher Pb(II) concentrations significantly reduces removal efficiency. This may be due to Pb(II) complexation or increased ionic competition, which alters electrostatic interactions between Pb(II) and the membrane.

For the quadratic effect, the negative coefficient of -15.02 feed concentration (A^2 = -15.02) indicates a nonlinear decrease in Pb(II) removal at higher concentrations. Beyond an optimal point, adsorption sites become overwhelmed, resulting in a sharp decline in removal efficiency. The strongest quadratic effect is observed for pressure (B^2 = -20.34), suggesting that after an optimal pressure level, further increases reduce efficiency. This could be due to membrane compaction, which alters pore size and permeability, negatively

affecting Pb(II) rejection. Lastly, the negative quadratic effect of pH (C^2 = -8.94) implies that while moderate pH improves removal, excessive pH shifts the equilibrium towards Pb(II) precipitation rather than membrane adsorption, reducing effectiveness.

Residual Analysis

Residual analysis is used to evaluate the distribution of the residuals, normality, and presence of outliers (Nobre & da Motta Singer, 2007). To further investigate four diagnostic graphs, predicted responses versus observed responses, normal probability plot, residuals versus predicted responses, and residuals versus run order were examined as depicted in Figure 6. Based on the plots analyzed, from Figure 7a, the predicted response versus observed response shows that the prediction deviation does not deviate far from the normal

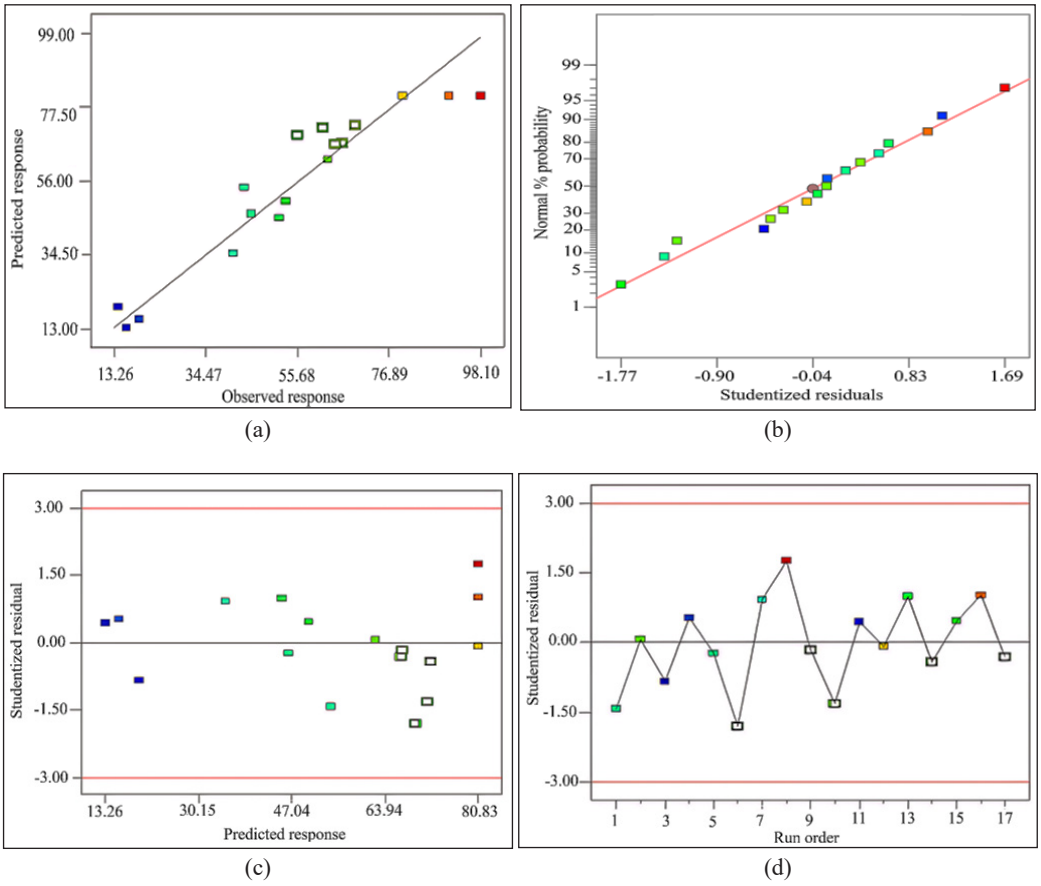


Figure 7. The residuals analysis diagnostic plots for (a) predicted against observed responses, (b) normal probability plot, (c) residuals against predicted responses, and (d) residuals against run order
Note. Red lines refer to the studentized residual limits

line. On top of that, similar plotting can be seen in Figure 7b, which shows the validity of the ANOVA results (Tiwari et al., 2017). Apart from that, the residual plot versus predicted response shows the variance homogeneity based on the distribution of residual plots randomly on both sides of the normal line, which denotes the consistency of the variance (Halim et al., 2021). In addition to that, the scattered residual points from diagnostic plots in Figure 7c exhibit no outliers; that is, the residual points are only distributed randomly within the upper and lower limit lines (Vining, 2010). Similarly, the residuals plot versus run order, as shown in Figure 7d, shows residual plot inside the upper and lower limit lines, and residual points were scattered randomly about the normal line with no discernible patterns, indicating that the residuals are independent of one another and hence obey the ANOVA assumptions. The ANOVA assumptions were met because of the residuals analysis, and it can be said that the regression model produced unbiased coefficient estimates with minimum variance.

Effect of Filtration Factors Based on the Response Surface Plots Analysis

The interaction between the independent factors and the responses was visualized using response surface plots, presented as both two-dimensional (2D) contour plots and three-dimensional (3D) surface plots, as shown in Figure 7. Based on the polynomial model from Equation 4, the 3D surface plots were generated by analyzing the combined influence of two factors at a time, with the third factor maintained at its mid-level value. The color gradients in the plots, ranging from red to green to blue, represent the highest, moderate, and lowest levels of Pb(II) ion rejection, respectively.

The CCD–RSM surfaces and contour plots (Figures 8a–d) show a single broad optimum for Pb(II) rejection (R1) as a quadratic function of feed concentration (A), pressure (B), and pH (C). Along the A–B plane (Fig. 8a and 8c), rejection rose when concentration increased from 60 to 135 ppm and pressure from 1.6 to 2.0 bar, reaching a predicted maximum of ~82% at A=135 ppm and B=2.0 bar, after which R1 declined towards the design-space edges. The initial increase reflected stronger electrostatic exclusion and favorable convective transport at moderate loading, whereas the downturn at higher A and/or B was consistent with concentration-polarization/fouling that diminished the effective driving force and apparent selectivity (Bowen & Mohammad, 1998; Donnan, 1995; Tsuru et al., 1991; Abdullah et al., 2018; Qadir et al., 2017). On the A–C plane (Figures 8b and 8d), rejection improved with increasing alkalinity and peaked at a pH of approximately 11 at a feed concentration of A = 135 ppm, attributable to the deprotonation of surface groups on the JPU/CDs and enhanced electrostatic repulsion of Pb(II). At lower pH, protonation screens charge and reduces Donnan exclusion (Elimelech & Childress, 1996; Marecka-Migacz et al., 2020). These interactions highlight that moderate concentration, moderate pressure, and mildly alkaline conditions acted synergistically, whereas pushing any factor

to extremes triggered polarization/compaction consequences (Amin et al., 2011; Jamalludin et al., 2021; Yaroshchuk, 2008).

RSM optimization predicted an optimum Pb(II) rejection of around 82% for the JPU/CDs (0.65 wt% CDs) at A = 135 ppm, B = 2 bar, and C = pH 11. Relative to pristine JPU (33% rejection), the incorporation of CDs elevated performance to the 80% class under these operating conditions, confirming the efficacy of the carbon-dot modification.

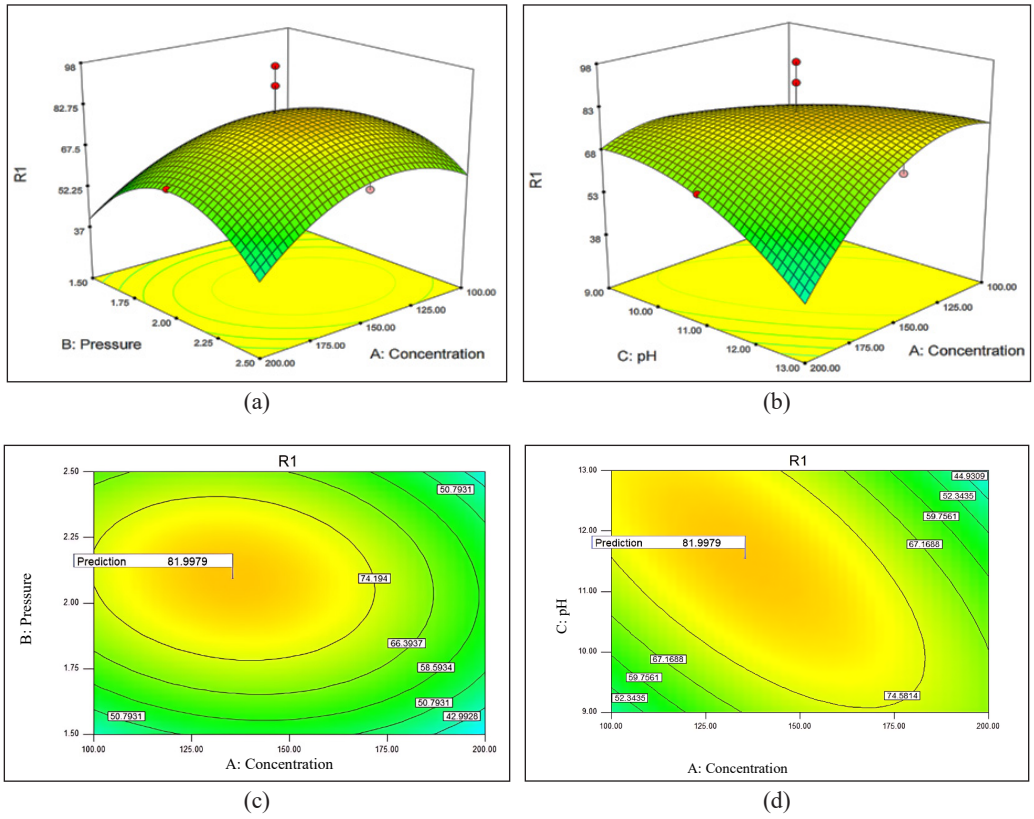


Figure 8. Response-surface (a, b) and contour (c, d) plots for Pb(II) rejection (R1) by JPU/CDs (0.65 wt% CDs): (a, c) concentration–pressure (A–B) interaction and (b, d) concentration–pH (A–C) interaction. A dome-shaped optimum occurs near A = 135 ppm, B = 2.0 bar, and C = pH 11 (predicted R1 = 82%)
Note. JPU = Jatropha polyurethane; CDs = Carbon dots; A = Concentration; B = Pressure; C = pH; CCD = Central composite design, RSM = Response surface methodology

Optimization and Model Validation

The optimization aims to maximize the Pb(II) ions rejection at the suggested conditions. Based on the results obtained, the optimum condition was proposed at 135 ppm, 2 bar, and at pH 11 to obtain the maximum Pb(II) ions rejection (82%), based on the desirability function of 0.8, which indicates the reasonable approximation of the prediction point.

However, in an actual filtration process, JPU/CDs 0.65 wt% replicated with the same optimum condition showed a 67% rejection with 18% error. Pb(II) rejection performance is fairly good compared to the others.

Table 5 compares the filtration performance of pristine JPU, JPU modified with 0.50 wt% GO, and JPU blended with 0.65 wt% CDs. In terms of water permeability, the CD-modified membrane shows a step-change with its preliminary water flux of 873 L/m².h, which is almost 4 times that of pristine JPU (223 L/m².h) and 1.7 times that of the GO composite (523 L/m².h). The likely reason is that the nano-sized CDs create additional porosity and hydrophilic nano-channels, reducing hydraulic resistance. By contrast, the GO membrane more than doubles the pristine flux (+135 %, from 223 to 523 L/m².h); its plate-like sheets introduce hydrophilic pathways but also increase tortuosity, so the permeability boost is significant but not as dramatic as with CDs (Harun et al., 2020).

For heavy metals removal applications, the GO composite delivers the highest selectivity of 71.6 % in the preliminary test for copper ions (Cu^{2+}) removal, rising to 82 ± 6% after optimization (Harun et al., 2022). In this study, the CD membrane targets Pb^{2+} and achieves 69% rejection in the preliminary run and 67 ± 18% under optimized conditions, which is roughly double the performance of pristine JPU (33%), but 15% below the GO membrane. Direct quantitative comparison is limited by the different target ions (Cu^{2+} vs. Pb^{2+}), yet the data still illustrate the contrasting strengths of each nanofiller.

With its much higher flux (767 L/m² h at optimum) and moderate Pb^{2+} rejection, the CD membrane offers the greatest treated-water throughput, which is advantageous for high-flow applications or a multi-stage polishing strategy. The GO membrane appeared to excel in rejection and is therefore better suited to single-stage operations for application with stringent heavy-metal limits. In summary, nanoparticle functionalization is essential for upgrading JPU membranes.

Table 5
Filtration performance of pristine jatropha polyurethane (JPU), JPU with graphene oxide (GO), and JPU with carbon dots (CDs)

	Unmodified pristine JPU	**Modified JPU with 0.50% wt GO**	**JPU/CDs with 0.65wt% CDs (this study)**
Preliminary water flux (L/m².h)	223[1]	523[1]	873
Preliminary heavy metal rejection (%)	33[1]	71.6[1]	69
Heavy metal rejection % and flux (L/m².h) at optimum conditions	33%[1] 223 L/m².h	87% (± 6% error)[2] Not mentioned[2]	67% (± 18% error) 767 L/m².h

Note. [1] Harun et al. (2020); [2]Harun et al. (2022)

Other studies on the implementation of RSM for nanomaterial-incorporated membranes in filtration have demonstrated varying levels of heavy metal rejection. For instance, an optimized metal oxide membrane incorporated with a nanocomposite achieved a copper ion (Cu(II)) rejection of 49% following RSM optimization (Zahed et al., 2019). Similarly, a polysulfone membrane embedded with cellulose nanofilters exhibited a chromium (VI) ion (Cr(VI)) rejection of 79.85% after optimization through RSM (Gasemloo et al., 2019). Additionally, JPU/GOs membrane filtration, when optimized via RSM, resulted in a Cu(II) rejection of 73.60% while adsorption-based optimization via RSM has demonstrated Pb(II) rejection efficiencies as high as 99% and 92% (Alipour et al., 2020; Harun et al., 2020; Zaferani et al., 2019). Other removal methods, such as cation exchange resin and chemical precipitation, also gave higher removal of Pb(II) at 99% and 97% (Chen et al., 2018; Lalmi et al., 2018). In short, the preliminary Pb(II) rejection performance of JPU/CD membranes appeared to be relatively lower than other reported methods. Nevertheless, despite the lower rejection efficiency compared to other established methods, JPU/CD membranes still offer potential advantages, such as low cost and sustainability for heavy metal removal, and require further optimization.

CONCLUSION

A regression model was developed based on the experimental design to optimize Pb(II) ion removal efficiency. Under the identified optimal conditions of 135 ppm feed concentration, 2 bar pressure, and pH 11, the model predicted an 82% removal rate. The analysis indicated that lower feed pressure and solution pH had a more pronounced effect on Pb(II) removal, whereas an increase in feed concentration initially enhanced removal efficiency. However, at excessively high feed concentrations, a decline in rejection efficiency was observed, likely due to concentration polarization effects. To refine the predictive accuracy of the model, a reduction approach was applied by eliminating statistically insignificant terms ($p > 0.05$). The resulting quadratic regression model demonstrated statistical significance, confirming its suitability for optimization predictions. These findings highlight the effectiveness of the developed model in optimizing Pb(II) removal and provide a foundation for further improvements in membrane-based heavy metal remediation strategies.

ACKNOWLEDGEMENTS

We would like to thank the Ministry of Higher Education, Malaysia, under the Fundamental Research Grant Scheme (grant code FRGS/1/2020/TK0/UPM/02/38, project code 03-01-20-2248FR) and the Initiative Putra Grant (GPI/2022/9720400) for their support of this work.

REFERENCES

Abdullah, N., Rahman, M. A., Othman, M. H. D., Jaafar, J., & Ismail, A. F. (2018). Membranes and membrane processes: Fundamentals. In A. Basile, S. Mozia, R. Molinari (Eds.), *Current trends and future developments on (bio-) membranes: Photocatalytic membranes and photocatalytic membrane reactors* (pp. 45–70). Elsevier. https://doi.org/10.1016/B978-0-12-813549-5.00002-5

Abu-Zurayk, R., Alnairat, N., Khalaf, A., Ibrahim, A. A. Halaweh, G. (2023). Cellulose acetate membranes: Fouling types and antifouling strategies — A brief review. *Processes*, *11*(2), 489. https://doi.org/10.3390/pr11020489

Alipour, A., Zarinabadi, S., Azimi, A., & Mirzaei, M. (2020). Adsorptive removal of Pb(II) ions from aqueous solutions by thiourea-functionalized magnetic ZnO/nanocellulose composite: Optimization by response surface methodology (RSM). *International Journal of Biological Macromolecules*, *151*, 124–135. https://doi.org/10.1016/j.ijbiomac.2020.02.109

Al-Rashdi, B. A. M., Johnson, D. J., & Hilal, N. (2013). Removal of heavy metal ions by nanofiltration. *Desalination*, *315*, 2–17. https://doi.org/10.1016/j.desal.2012.05.022

Amin, M. M., Hafez, A. I., Shaaban, A. F., Abdelmonem, N. M., & Hanafy, M. (2011). *Removal of reactive dyes from dye-house effluent using nanofiltration membrane*. ResearchGate. https://www.researchgate.net/publication/279297599_Removal_of_Reactive_Dyes_from_Dye-house_Effluent_Using_Nanofiltration_Membrane

Behera, S. K., Meena, H., Chakraborty, S., & Meikap, B. C. (2018). Application of response surface methodology (RSM) for optimization of leaching parameters for ash reduction from low-grade coal. *International Journal of Mining Science and Technology*, *28*(4), 621–629. https://doi.org/10.1016/j.ijmst.2018.04.014

Bowen, W. R., & Mohammad, A. W. (1998). Characterization and prediction of nanofiltration membrane performance — A general assessment. *Chemical Engineering Research and Design*, *76*, 885–893. https://doi.org/10.1205/026387698525685

Chee, D. N. A., Aziz, F., Mohamed Amin, M. A., & Ismail, A. F. (2021). Copper adsorption on ZIF-8/alumina hollow fiber membrane: A response surface methodology analysis. *Arabian Journal for Science and Engineering*, *46*, 6775–6786. https://doi.org/10.1007/s13369-021-05636-1

Chen, Q., Yao, Y., Li, X., Lu, J., Zhou, J., & Huang, Z. (2018). Comparison of heavy metal removals from aqueous solutions by chemical precipitation and characteristics of precipitates. *Journal of Water Process Engineering*, *26*, 289–300. https://doi.org/10.1016/j.jwpe.2018.11.003

Cruz, R., Nisar, M., Palza, H., Yazdani-Pedram, M., Aguilar-Bolados, H., & Quijada, R. (2023). Development of bio degradable nanocomposites based on PLA and functionalized graphene oxide. *Polymer Testing*, *124*, 108066. https://doi.org/10.1016/j.polymertesting.2023.108066

Deshmukh, S. S., & Childress, A. E. (2001). Zeta potential of commercial RO membranes: Influence of source water type and chemistry. *Desalination*, *140*(1), 87–95. https://doi.org/10.1016/S0011-9164(01)00357-5

Donnan, F. G. (1995). Theory of membrane equilibria and membrane potentials in the presence of non-dialysing electrolytes. A contribution to physical-chemical physiology. *Journal of Membrane Science*, *100*(1), 45–55. https://doi.org/10.1016/0376-7388(94)00297-C

Duclos-Orsello, C., Kelly, W., Grant, D. C., Zahka, J., & Thom, V. (2004). Neutral adsorptive capture of particles by membranes: Network modeling near the membrane isoelectric point. *Journal of Membrane Science*, *237*(1–2), 167–180. https://doi.org/10.1016/j.memsci.2004.02.030

Elimelech, M., & Childress, A. E. (1996). *Zeta potential of reverse osmosis membranes: Implications for membrane performance*. National Technical Reports Library. https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/PB97127450.xhtml

Emamjomeh, M. M., Torabi, H., Mousazadeh, M., Alijani, M. H., & Gohari, F. (2019). Impact of independent and non-independent parameters on various elements' rejection by nanofiltration employed in groundwater treatment. *Applied Water Science*, *9*, 71. https://doi.org/10.1007/s13201-019-0949-1

Fang, H., Wang, X., Xia, D., Zhu, J., Yu, W., Su, Y., Zeng, J., Zhang, Y., Lin, X., Lei, Y., & Qiu, J. (2022). Improvement of ecological risk considering heavy metal in soil and groundwater surrounding electroplating factories. *Processes*, *10*(7), 1267. https://doi.org/10.3390/pr10071267

Feng, Z., Adolfsson, K. H., Xu, Y., Fang, H., Hakkarainen, M., & Wu, M. (2021). Carbon dot/polymer nanocomposites: From green synthesis to energy, environmental and biomedical applications. *Sustainable Materials and Technologies*, *29*, e00304. https://doi.org/10.1016/j.susmat.2021.e00304

Gao, F. (2019). An overview of surface-functionalized magnetic nanoparticles: Preparation and application for wastewater treatment. *ChemistrySelect*, *4*(22), 6805–6811. https://doi.org/10.1002/slct.201900701

Gasemloo, S., Khosravi, M., Sohrabi, M. R., Dastmalchi, S., & Gharbani, P. (2019). Response surface methodology (RSM) modeling to improve removal of Cr (VI) ions from tannery wastewater using sulfated carboxymethyl cellulose nanofilter. *Journal of Cleaner Production*, *208*, 736–742. https://doi.org/10.1016/j.jclepro.2018.10.177

Halim, N. A. A., Abidin, Z. Z., Siajam, S. I., Hean, C. G., & Harun, M. R. (2021). Optimization studies and compositional analysis of subcritical water extraction of essential oil from *Citrus hystrix* DC. leaves. *The Journal of Supercritical Fluids*, *178*, 105384. https://doi.org/10.1016/j.supflu.2021.105384

Harun, N. H., Abidin, Z. Z., Abdullah, A. H., & Othaman, R. (2020). Sustainable *Jatropha* oil-based membrane with graphene oxide for potential application in Cu(II) ion removal from aqueous solution. *Processes*, *8*(2), 230. https://doi.org/10.3390/pr8020230

Harun, N. H., Abidin, Z. Z., Majid, U. A., Hamid, M. R. A., Abdullah, A. H., Othaman, R., & Harun, M. Y. (2022). Adopting sustainable *Jatropha* oil bio-based polymer membranes as alternatives for environmental remediation. *Polymers*, *14*(16), 3325. https://doi.org/10.3390/polym14163325

Hu, Q., Yuan, Y., Wu, Z., Lu, H., Li, N., & Zhang, H. (2023). The effect of surficial function groups on the anti-fouling and anti-scaling performance of thin-film composite reverse osmosis membranes. *Journal of Membrane Science*, *668*, 121276. https://doi.org/10.1016/j.memsci.2022.121276

Idress, H., Zaidi, S. Z. J., Sabir, A., Shafiq, M., Khan, R. U., Harito, C., Hassan, S., & Walsh, F. C. (2021). Cellulose acetate based complexation-NF membranes for the removal of Pb(II) from waste water. *Scientific Reports*, *11*, 1806. https://doi.org/10.1038/s41598-020-80384-0

Issa, M. A., Abidin, Z. Z., Pudza, M. Y., & Zentou, H. (2020). Efficient removal of Cu(II) from aqueous systems using enhanced quantum yield nitrogen-doped carbon nanodots. *RSC Advances*, *10*(25), 14979–14990. https://doi.org/10.1039/D0RA02276D

Issa, M. A., Abidin, Z. Z., Sobri, S., Rashid, S. A., Mahdi, M. A., & Ibrahim, N. A. (2020). Fluorescent recognition of Fe^{3+} in acidic environment by enhanced-quantum yield N-doped carbon dots: Optimization of variables using central composite design. *Scientific Reports*, *10*, 11710. https://doi.org/10.1038/s41598-020-68390-8

Jamalludin, M. R., Ahmad Termizi, S. N. A., Khor, C. Y., Hubadillah, S. K., Nawi, M. A. M., Ishak, M. I., & Rosli, M. U. (2021). The modified phase inversion and sintering technique for fabrication of ceramic membrane: Influence of kaolin composition and drying temperature. In *AIP Conference Proceedings* (Vol. 2339, No. 1, p. 020186). AIP Publishing. https://doi.org/10.1063/5.0044560

Junker, M. A., Regenspurg, J. A., Valdes Rivera, C. I., te Brinke, E., & de Vos, W. M. (2023). Effects of feed solution pH on polyelectrolyte multilayer nanofiltration membranes. *ACS Applied Polymer Materials*, *5*(1), 355–369. https://doi.org/10.1021/acsapm.2c01542

Keçili, R., Büyüktiryaki, S., & Hussain, C. M. (2020). Membrane applications of nanomaterials. In C. M. Hussain (Ed.), *Handbook of nanomaterials in analytical chemistry: Modern trends in analysis* (pp. 159–182). Elsevier. https://doi.org/10.1016/B978-0-12-816699-4.00007-4

Khamis, F., Hegab, H. M., Banat, F., Arafat, H. A., & Hasan, S. W. (2023). Development of sustainable pH-responsive adsorptive modified mangrove-based polylactic acid ultrafiltration membrane for the removal of heavy metals from aqueous solution. *Chemical Engineering Journal*, *474*, 145471. https://doi.org/10.1016/j.cej.2023.145471

Khdary, N. H., Almuarqab, B. T., & El Enany, G. (2023). Nanoparticle-embedded polymers and their applications: A review. *Membranes*, *13*(5), 537. https://doi.org/10.3390/membranes13050537

Lalmi, A., Bouhidel, K.-E., Sahraoui, B., & el Houda Anfif, C. (2018). Removal of lead from polluted waters using ion exchange resin with $Ca(NO_3)_2$ for elution. *Hydrometallurgy*, *178*, 287–293. https://doi.org/10.1016/j.hydromet.2018.05.009

Low, K. H., Koki, I. B., Juahir, H., Azid, A., Behkami, S., Ikram, R., Mohammed, H. A., & Md. Zain, S. (2016). Evaluation of water quality variation in lakes, rivers, and ex-mining ponds in Malaysia (review). *Desalination and Water Treatment*, *57*(58), 28215–28239. https://doi.org/10.1080/19443994.2016.1185382

Mamba, F. B., Mbuli, B. S., & Ramontja, J. (2021). Recent advances in biopolymeric membranes towards the removal of emerging organic pollutants from water. *Membranes*, *11*(11), 798. https://doi.org/10.3390/membranes11110798

Marecka-Migacz, A., Mitkowski, P. T., Nędzarek, A., Różański, J., & Szaferski, W. (2020). Effect of pH on total volume membrane charge density in the nanofiltration of aqueous solutions of nitrate salts of heavy metals. *Membranes*, *10*(9), 235. https://doi.org/10.3390/membranes10090235

Marlina., Farida, M., & Mustanir. (2017). Synthesis and characterization of polyurethane membrane from nyamplung seed oils (*Calophyllum inophillum*). *Asian Journal of Chemistry*, *29*(9), 1912–1916. https://doi.org/10.14233/ajchem.2017.20568

Morales-Jiménez, M., Palacio, D. A., Palencia, M., Meléndrez, M. F., & Rivas, B. L. (2023). Bio-based polymeric membranes: Development and environmental applications. *Membranes*, *13*(7), 625. https://doi.org/10.3390/membranes13070625

Nobre, J. S., & da Motta Singer, J. (2007). Residual analysis for linear mixed models. *Biometrical Journal*, *49*(6), 863–875. https://doi.org/10.1002/bimj.200610341

Petrović, Z. S. (2008). Polyurethanes from vegetable oils. *Polymer Reviews*, *48*(1), 109–155. https://doi.org/10.1080/15583720701834224

Poerio, T., Denisi, T., Mazzei, R., Bazzarelli, F., Piacentini, E., Giorno, L., & Curcio, E. (2022). Identification of fouling mechanisms in cross-flow microfiltration of olive-mills wastewater. *Journal of Water Process Engineering*, *49*, 103058. https://doi.org/10.1016/j.jwpe.2022.103058

Qadir, D., Mukhtar, H. B., & Keong, L. K. (2017). Rejection of divalent ions in commercial tubular membranes: Effect of feed concentration and anion type. *Sustainable Environment Research*, *27*(2), 103–106. https://doi.org/10.1016/j.serj.2016.12.002

Sgreccia, E., Gonzalez, F. S. G., Prosposito, P., Burratti, L., Sisani, M., Bastianini, M., Knauth, P., & Di Vona, M. L. (2024). Heavy metal detection and removal by composite carbon quantum dots/ionomer membranes. *Membranes*, *14*(6), 134. https://doi.org/10.3390/membranes14060134

Shabbir, H., Csapó, E., & Wojnicki, M. (2023). Carbon quantum dots: The role of surface functional groups and proposed mechanisms for metal ion sensing. *Inorganics*, *11*(6), 262. https://doi.org/10.3390/inorganics11060262

Shi, X., Tal, G., Hankins, N. P., & Gitis, V. (2014). Fouling and cleaning of ultrafiltration membranes: A review. *Journal of Water Process Engineering*, *1*, 121–138. https://doi.org/10.1016/j.jwpe.2014.04.003

Stade, S., Kallioinen, M., Mikkola, A., Tuuva, T., & Mänttäri, M. (2013). Reversible and irreversible compaction of ultrafiltration membranes. *Separation and Purification Technology*, *118*, 127–134. https://doi.org/10.1016/j.seppur.2013.06.039

Tiwari, A., Pal, D., & Sahu, O. (2017). Recovery of copper from synthetic solution by efficient technology: Membrane separation with response surface methodology. *Resource-Efficient Technologies*, *3*(1), 37–45. https://doi.org/10.1016/j.reffit.2016.12.008

Tsuru, T., Urairi, M., Nakao, S.-I., & Kimura, S. (1991). Reverse osmosis of single and mixed electrolytes with charged membranes: Experiment and analysis. *Journal of Chemical Engineering of Japan*, *24*(4), 518–524.

van der Kuijp, T. J., Huang, L., & Cherry, C. R. (2013). Health hazards of China's lead-acid battery industry: A review of its market drivers, production processes, and health impacts. *Environmental Health*, *12*, 61. https://doi.org/10.1186/1476-069X-12-61

Vining, G. (2010). Technical advice: Residual plots to check assumptions. *Quality Engineering*, *23*(1), 105–110. https://doi.org/10.1080/08982112.2011.535696

Wan Yaacob, W. Z., Mohd Pauzi, N. S., & Abdul Mutalib, H. (2009). Acid mine drainage and heavy metals contamination at abandoned and active mine sites in Pahang. *Bulletin of the Geological Society of Malaysia*, *55*, 15–20. https://doi.org/10.7186/bgsm55200903

Xia, C., Ye, H., Wu, Y., AL. Garalleh, H., Garaleh, M., Sharma, A., & Pugazhendhi, A. (2023). Nanofibrous/biopolymeric membrane a sustainable approach to remove organic micropollutants: A review. *Chemosphere*, *314*, 137663. https://doi.org/10.1016/j.chemosphere.2022.137663

Yaroshchuk, A. E. (2008). Negative rejection of ions in pressure-driven membrane processes. *Advances in Colloid and Interface Science*, *139*(1–2), 150–173. https://doi.org/10.1016/j.cis.2008.01.004

Zaferani, S. P. G., Emami, M. R. S., Amiri, M. K., & Binaeian, E. (2019). Optimization of the removal Pb(II) and its Gibbs free energy by thiosemicarbazide modified chitosan using RSM and ANN modeling. *International Journal of Biological Macromolecules*, *139*, 307–319. https://doi.org/10.1016/j.ijbiomac.2019.07.208

Zahed, S. S. H., Khanlari, S., & Mohammadi, T. (2019). Hydrous metal oxide incorporated polyacrylonitrile-based nanocomposite membranes for Cu(II) ions removal. *Separation and Purification Technology*, *213*, 151–161. https://doi.org/10.1016/j.seppur.2018.12.027

Zularisam, A. W., Ismail, A. F., Salim, M. R., Sakinah, M., & Hiroaki, O. (2007). Fabrication, fouling and foulant analyses of asymmetric polysulfone (PSF) ultrafiltration membrane fouled with natural organic matter (NOM) source waters. *Journal of Membrane Science*, *299*(1–2), 97–113. https://doi.org/10.1016/j.memsci.2007.04.030

Pertanika J. Sci. & Technol. 33 (6): 2797 - 2832 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

RGB vs. HSV for Kitchen Fire Detection with YOLOv5

Norisza Dalila Ismail, Rizauddin Ramli* and Mohd Nizam Ab Rahman

Department of Mechanical and Manufacturing Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

ABSTRACT

Kitchen fires pose a significant challenge and threat to people and the environment. Prompt response and accurate classification of fire occurrences are crucial to ensure safety and reduce potential property damage. This study addresses the need for effective fire detection technologies by evaluating the performance of the You Only Look Once version 5 medium (YOLOv5m) model using both hue, saturation, and value (HSV) and visible light (red, green, and blue [RGB]) color spaces. To reduce false positives, the experiment was expanded by including background images in the training dataset. Two kitchen fire datasets, one in RGB and one in HSV, were used to train and evaluate the model. Based on the results, HSV color space offers higher recall and precision for fire-on-pan detection, achieving 0.882 and 0.931, outperforming RGB. The best overall mean performance was in the first experiment, RGB without background images, resulting in the highest mean average precision (mAP)@0.5:0.95 score of 0.651. This performance was limited by lower recall and precision compared to HSV in specific fire scenarios. A key limitation of this study is its focus on kitchen environments, for which the findings may not be directly generalizable to other fire scenarios with different environmental conditions or fire characteristics. Furthermore, the study is limited to the YOLOv5m architecture, where other detection models may yield different results. In terms of kitchen fire detection, this study provides a comprehensive comparison of the RGB and HSV color spaces, offering insights into their benefits and drawbacks. This research shows that the HSV color space is useful in certain fire detection scenarios, and that combining the two color spaces yields an improved detection model for real-time applications.

ARTICLE INFO

Article history:
Received: 06 December 2024
Accepted: 03 July 2025
Published: 29 October 2025

DOI: https://doi.org/10.47836/pjst.33.6.17

E-mail addresses:
i.norisza@gmail.com (Norisza Dalila Ismail)
rizauddin@ukm.edu.my (Rizauddin Ramli)
mnizam@ukm.edu.my (Mohd Nizam Ab Rahman)
* Corresponding author

Keywords: Deep learning, HSV, kitchen fire detection, object detection, RGB, YOLOv5

INTRODUCTION

A fire starting from the kitchen extended farther than the cooking space and could easily burn combustibles kept in other areas of the house. Flashover may occur if the fire

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

spreads swiftly. The appearance of flashover, which is a crucial factor in catastrophic fires, is typically interpreted as a sign of untenable conditions (Rasbash, 1991). The fire will spread more quickly and could endanger the entire building. This highlights the critical need for more sustainable and effective methods to prevent fires. Consequently, to address the fire risk issue, an innovative fire detection system must be developed.

Fire detection systems are designed to detect fires early, providing sufficient time to extinguish them and ensure the safe evacuation of residents. Early detection may protect the safety of emergency response personnel, reduce property losses, and minimize operational downtime since control measures are initiated (Graham & Roberts, 2000). Existing fire detection systems can include detectors for smoke, heat, or flame, which are connected to a fire alarm system that triggers an alarm to alert people to potential danger.

However, these systems have difficulties in accurately detecting fires in dynamic environments like kitchens, where smoke and steam are frequent byproducts of cooking activities (Choi et al., 2023; Milke & Zevotek, 2016). Due to the complex conditions in kitchen settings, such as variable lighting, reflective surfaces, and background clutter, it is becoming increasingly challenging for conventional sensors to differentiate between real fire occurrences. Therefore, there is a pressing need for more advanced and reliable fire detection systems capable of detecting genuine fire events early in typical kitchen environments. Transitions of various innovative methods have been introduced to improve fire detection systems, such as vision-based fire detection (Bu & Gharajeh, 2019; Dilshad et al., 2023). Vision-based systems use image texture and color to identify flames. RGB and HSV are two of the most used color spaces in fire detection because they facilitate the easy distinction between fire and other elements in a kitchen environment. Giwa and Benkrid (2024) suggested systems vary in complexity; to recognize distinguishing fire characteristics in a video series, including color, velocity, temporal variations, spatial variations, and dynamic texture, they use a variety of image processing, machine learning techniques, and convolutional neural network (CNN) structures.

The RGB color space is mostly used in fire detection, which presents a significant issue because it can be extremely sensitive to changes in illumination and reflections, resulting in false alarms (Bakri et al., 2018). On the other hand, the HSV color space is more resilient to changes in illumination since it distinguishes between color intensity and chromatic content. However, it is not completely error-proof (Mutar, 2019).

In machine learning, object detection is the process of identifying and categorizing objects that meet predetermined criteria. Examples include drawing boxes around the objects and labeling them with their assigned class. Modern object identification techniques like You Only Look Once (YOLO) have drawn much attention lately because of their dependability, speed, and accuracy (Redmon et al., 2015). Numerous studies and real-world applications have consistently demonstrated the efficiency of this novel approach, which has shown great promise across a range of applications.

YOLO was built on a CNN that predicts several bounding boxes at once and assigns class probabilities to each box in an image using a single neural network and a single forward propagation process. The evolution of YOLO architecture with the development of numerous iterations, including You Only Look Once version 1 (YOLOv1) to You Only Look Once version 11 (YOLOv11), which are YOLOv1 (Redmon et al., 2015), You Only Look Once version 2 (YOLOv2, Redmon & Farhadi, 2017), You Only Look Once version 3 (YOLOv3, Redmon & Farhadi, 2018), You Only Look Once version 4 (YOLOv4, Bochkovskiy et al., 2020), You Only Look Once version 5 (YOLOv5, Jocher, n.d.b), You Only Look Once version 6 (YOLOv6, Li et al., 2022), You Only Look Once version 7 (YOLOv7, C.-Y. Wang et al., 2022), You Only Look Once version 8 (YOLOv8, Jocher, n.d.a), You Only Look Once version 9 (YOLOv9, C.-Y. Wang et al., 2024), You Only Look Once version 10 (YOLOv10, A. Wang et al., 2024), and the latest in the year 2024 is YOLOv11 from Ultralytics. This version represents a significant improvement over the previous versions. Still, the official version exhibits some instability due to network-related updates. In this study, the YOLOv5 version (Jocher, n.d.b) was selected due to its network structure, which is more efficient than that of YOLOv8 (Geng et al., 2024), resulting in a smaller model size and improved deployment and operating efficiency. Furthermore, YOLOv5 enhances object detection performance by employing multiple optimization techniques and a less computationally intensive backbone, making it an excellent choice for real-time applications (Geng et al., 2024). With a pre-trained model and an intuitive application programming interface (API), YOLOv5 is a convenient and user-friendly choice.

Existing fire detection methods often struggle with the diverse and dynamic conditions prevalent in kitchen environments, including variations in lighting, the presence of steam and smoke from cooking, and the wide range of background objects. This paper aims to address these limitations by investigating the efficacy of a YOLOv5-based fire detection system that uses both RGB and HSV color spaces for early kitchen fire detection. Specifically, the primary objective of this study is to evaluate the performance of RGB and HSV color spaces when integrated with the YOLOv5 object detection model for detecting early-stage kitchen fires.

This work focuses on detecting visible kitchen fires in images taken in realistic kitchen environments. Although the initial plan included addressing the problem of early fire detection, the available dataset obtained so far contains mainly active fire images. For this reason, pre-flame states, including those for smoke-only and smoldering conditions, are not considered in this paper. The comparative study presented in this article focuses on the impact of color space representations (RGB versus HSV) on YOLOv5m's object detection performance in cluttered kitchen environments.

Two distinct datasets were created: one containing images in the RGB color space and the other in the HSV color space. This study systematically compares the performance of

these two color spaces within the YOLOv5 framework to identify an optimal approach for accurately distinguishing fire from non-fire events in real-time kitchen scenarios.

RELATED WORKS

YOLOv5 has achieved significant gains in both accuracy and speed for real-time applications, and numerous works have focused on improving its architecture for fire and smoke detection. However, most of these models operate primarily in the RGB color space, without assessing how color representation may influence detection performance under challenging indoor environments, such as kitchens. Table 1 shows advances in object detection models, particularly in architectural improvements for vision-based fire detection. These include integrated channel attention and transformer mechanisms from Liang et al. (2023), while Geng et al. (2024) presented YOLOFM, which integrates the FocalNext block, a quantization-aware hybrid attention-reparameterized feature pyramid network (QAHARep-FPN), and the Focal-scaled intersection over union (SIoU) loss function.

Another project has been developed, which involved You Only Look Once version 5 small (YOLOv5s) backbone replacement with EfficientNetV2 and introduction of a Smooth L1 loss function (S. Wang & Wang, 2023). Enhancement of the YOLOv5 framework by D. Zhang and Chen (2024), which incorporates the FasterNet network into its backbone, while the neck network uses Ghost-Shuffle Convolution (GSConv). They also introduced a one-time aggregation cross-stage partial network module, a one-shot aggregation network-based ghost shuffle cross-stage partial (VoV-GSCP) in the modified model.

A modified version of the YOLOv5m with real-time auditory notifications was introduced by Abdusalomov et al. (2022), which was designed to assist visually impaired individuals, especially during a fire. Kobeissi and Boulos (2023) explored the use of a combination of YOLOv5 and MiDaS depth estimators to enhance indoor fire detection.

However, these enhancements often remain bound to the RGB color domain, potentially overlooking the benefits offered by alternative color spaces. There is a distinct dearth of targeted research addressing fire detection specifically inside kitchen contexts, despite strong evidence suggesting that a significant number of fires originate in kitchens (Chow & Liu, 2016; J. Liu et al., 2019; Ngai et al., 2022; Spearpoint & Hopkin, 2020; Zulkiffli, 2018). Although a great deal of research has been conducted on fire detection devices, relatively little has focused on the characteristics and challenges of kitchen fires. This gap in the literature underscores the need for further research into fire detection techniques specifically designed for kitchens to enhance early warning systems in high-risk environments.

Table 1
Summary of related work

Authors	Models' enhancements	Dataset	Color space used	Performance metrics	Findings	
D. Zhang and Chen (2024)		The backbone network integrates YOLOv5 with the FasterNet network. The neck network uses GSConv and VoV-GSCSP.	Public dataset	RGB	Average accuracy = 98.3%, reduction in memory usage = 31.4%, increase in detection speed = 13%.	The model is lightweight, maintaining accuracy. Reduce model memory usage, model parameters, and computational costs. Improve detection speed.
Geng et al. (2024)	Improved YOLOv5n with QARepNeXt for feature extraction, Integration of QAHARep-FPN, and Focal-SIoU loss function.	FM-VOC Dataset18644	RGB	mAP50-95 = 7.9% improvement.	Improved feature extraction for better fire and smoke detection.	
Liang et al. (2023)	Channel attention and transformer in YOLOv5s.	Self-built internet fire image	RGB	14% mAP improvement and 68 FPS.	Enhanced detection speed and accuracy for fire and smoke.	
S. Wang and Wang (2023)	YOLOv5s backbone replaced with EfficientNetV2 and SIoU loss.	PASCAL VOC2007	RGB	mAP = 20% improvement, 15% recall improvement.	Lightweight and real-time fire detection with improved accuracy.	
Kobeissi and Boulos (2023)	YOLOv5 and MiDaS depth-mapper, depth estimator.	Fire and smoke dataset 2) Fire and gun on Kaggle	RGB + Depth	98.26% accuracy was achieved in the larger MiDaS version.	Locating and detecting indoor fires with the use of a single camera.	

Table 1 *(continue)*

Authors	Models' enhancements	Dataset	Color space used	Performance metrics	Findings
Abdusalomov et al. (2022)	YOLOv5m with real-time auditory notifications for the visually impaired.	Publicly available image datasets and Google images	RGB	High-speed, high-accuracy fire detection in indoor scenes.	Real-time fire detection for indoor environments.

Note. YOLOv5 = You Only Look Once version 5; GSConv = Ghost-shuffle convolution; VoV-GSCSP = One-shot aggregation network-based ghost shuffle cross stage partial; RGB = Red, green, and blue; YOLOv5n = You Only Look Once version 5 nano; QARepNeXt = Quantization-aware re-parameterized network with aggregated transformations; QAHARep-FPN = Quantization-aware hybrid attention re-parameterized feature pyramid network; Focal-SIoU = Focal scaled intersection over union loss function; FM-VOC Dataset 18644 = Fire and smoke visual object classes dataset with 18,644 images; mAP50-95 = Mean average precision at IoU thresholds from 0.50 to 0.95; YOLOv5s = You Only Look Once version 5 small; mAP = Mean average precision; FPS = Frame per second; EfficientNetV2 = Efficient convolutional neural network version 2; SIoU = Scaled intersection over union; PASCAL VOC2007 = Pattern analysis, statistical modelling and computational learning – Visual object classes challenge 2007; MiDaS = Mixed depth and scale

In this study, three significant key contributions are presented:

1. A comparative analysis of RGB and HSV color spaces specifically for kitchen fire detection, an area that has received limited attention.
2. The use of YOLOv5m, a state-of-the-art object detection model, for this task, which offers a balance between speed and accuracy.
3. An evaluation of the impact of background complexity on fire detection performance, providing insights into the robustness of the proposed approach in real-world scenarios.

Although previous work has examined RGB images and various deep learning models for fire detection, little research has investigated discriminative analysis between the RGB and HSV color spaces within a YOLOv5 architecture. To the best of our knowledge, there is no systematic work evaluating RGB vs. HSV color spaces for fire detection in kitchens using YOLOv5m, particularly about background variation and clutter, a crucial issue in real-world kitchen environments.

This study addresses these limitations by evaluating and comparing the performance of RGB and HSV for the YOLOv5m object detection model in kitchen fire scenarios. With this motivation, the research is directed by the following questions:

RQ1: How does the HSV color space perform with respect to detection (i.e., precision, recall, and mAP) in comparison with the RGB color space when YOLOv5m is used for kitchen fire detection?

RQ2: What is the impact of including realistic kitchen background images in the training dataset on the detection performance and the number of false positives of a YOLOv5m-based kitchen fire detection system?

These questions were explored through a set of experiments on RGB and HSV datasets, with identical model settings for fair comparisons.

PROPOSED METHODS

Research Overview

The suggested method for creating an indoor kitchen fire detection model is shown in Figure 1. A thorough method for creating a fire detection model is described in the suggested flowchart. The process starts with image database acquisition, followed by data pre-processing, which involves several steps such as data cleaning, cropping, resizing, removing background, and annotation. To verify the model's applicability, the pre-processed data is then split into three datasets, which are training, validation, and testing. Using the training and validation datasets, respectively, the model is trained and validated during the training phase. Next, the testing dataset is used to evaluate the refined fire detection model's performance on an unknown dataset. The model's recall rate, precision rate, mAP@0.5, and mAP@0.5-95 are all evaluated as part of the evaluation process. With the potential to significantly reduce environmental fires, this suggested methodology offers a methodical and efficient approach to building kitchen fire-detection models.

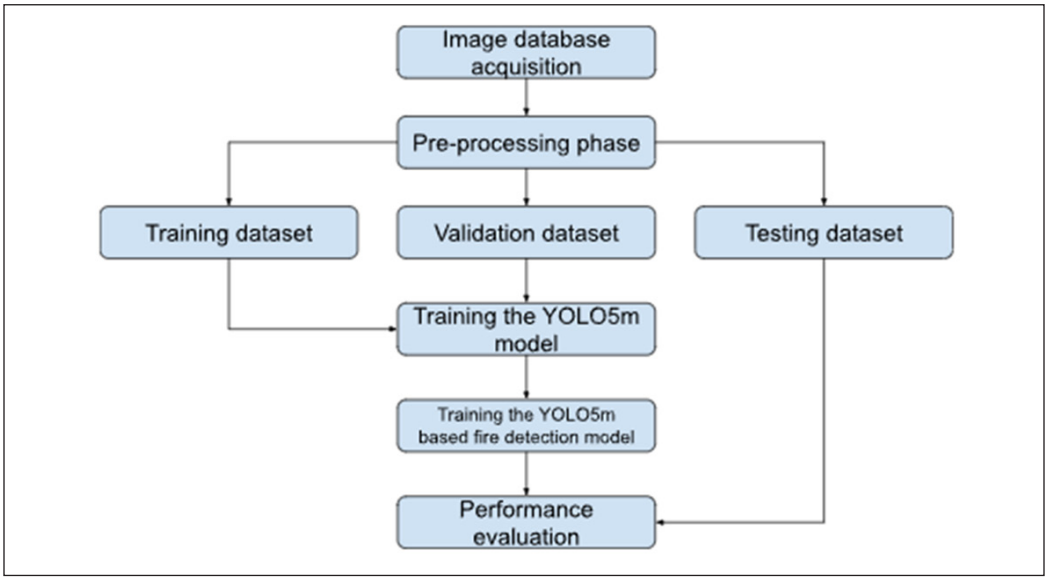


Figure 1. Flowchart of the proposed method
Note. YOLO5m = You Only Look Once version 5 medium

Object Detection Model

One object detection algorithm technique that is popular for its reliability, simplicity, and accuracy (Johnston et al., 2022). It is called YOLOv5. The YOLOv5 (Jocher, n.d.b) was released by Ultralytics and consists of three main components: the backbone, neck, and head.

The backbone depends on the CSP-Darknet53 convolutional network and employs the cross-stage partial (CSP) strategy to expedite information flow while alleviating issues linked to redundant and vanishing gradients (S. Liu et al., 2022; Wu et al., 2023). A version of the spatial pyramid pooling (SPP) is incorporated into the neck of the YOLOv5 model, and the Bottle-NeckCSP is integrated into the Path Aggregation Network (PANet) (Jocher, n.d.a). A combination of these techniques enhances the receptive field, allowing the network to retain speed while separating important context features. The CSPNet strategy (Wu et al., 2023) improved PANet in YOLOv4 as a feature pyramid network, providing better pixel-level localization accuracy in YOLOv5. The neck of architecture is crucial for managing object scaling and enabling the model to perform remarkably well on unseen data.

The head of YOLOv5 consists of three convolutional layers, similar to its predecessors (Wu et al., 2023). These layers, which differ slightly from the preceding versions in the computation of target coordinates for bounding boxes, predict the coordinates of bounding boxes, score, and object classification (Kim et al., 2023; Wu et al., 2023).

According to Johnston et al. (2022), there are four different levels of neural network models in YOLOv5, from simplest to most complex including YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. With the help of an effective inference strategy and a lightweight network architecture, YOLOv5s is a single-stage detector that enables quick and precise object recognition (Vats & Anastasiu, 2023). Besides, YOLOv5l and YOLOv5x offer higher accuracy due to their deeper and wider architecture, but require significantly more computational resources, which can be a limitation for deployment on hardware with constrained processing power. YOLOv5m represents a balanced compromise between these extremes, offering improved detection precision over YOLOv5s while maintaining a reasonable inference speed and manageable resource consumption. Given the demand for quick, medium-level performance, and high precision without overwhelming computational cost, YOLOv5m is selected as the model for this paper.

Performance Metrics

In this study, mAP, precision, recall, F1-score, and latency are selected as the primary evaluation metrics to assess the fire detection model's performance. These metrics are widely used in object detection tasks because they capture both localization and classification performance. Precision indicates the proportion of correctly identified fire instances among all instances predicted as fire, while recall measures the model's ability to

detect all actual fire instances. mAP, particularly at intersection over union (IoU) thresholds such as mAP@0.5 and mAP@0.5:0.95, provides a comprehensive summary by integrating precision and recall across multiple confidence levels and IoU thresholds. F1-score, the harmonic means of precision and recall, balances these two metrics, offering a single measure of a model's overall detection performance.

Since this study is limited to fire as a class, the computed mAP approximates the average precision (AP) across all fires. To compute the AP, the IoU must be provided first. The overlap between the ground-truth and predicted bounding boxes is measured using a metric called IoU. It is computed as the ratio of the intersection area to the union area of the two bounding boxes. This metric is frequently used to evaluate object recognition models, as it assesses how well the model can locate and identify objects in an image. A threshold value is compared with the IoU value for the anticipated and ground-truth bounding boxes to determine whether a detection is correct or incorrect. This study selected a threshold value of 0.5, which is commonly used in evaluating object detection models.

The link between the IoU value and the threshold value must be considered to assess an object detection model's performance properly. The terms listed below are introduced specifically to explain this relationship: For an IoU of 0.5 or greater, the results are considered true positives (TP); meanwhile, an IoU below 0.5 is classified as a false positive (FP), and a false negative (FN). TP denotes a ground truth bounding box that has been correctly detected, while FP denotes an item that has been incorrectly detected. FN denotes that the model cannot identify an object within the image. Since a true negative (TN) implies that there is no targeted object in the detected image, it was not considered while calculating performance (Padilla et al., 2020).

TNs are a crucial component of a model's performance, but they are usually excluded from the mAP computation for object detection since mAP is intended to assess the model's object recognition capabilities rather than its detection accuracy. Thus, TN is not relevant for this study, given that every fire image has previously been labelled during the pre-processing stage. Consequently, the evaluation performance will involve TP, FP, and FN to calculate the model's accuracy and reliability using precision (P) and recall (R). P will calculate the proportion of correct predictions to the model's accuracy in the detection process. The ratio of accurate positive forecasts to all ground facts is known as R. The P and R formulas, as indicated by Equations 1 and 2, respectively.

$$\text{Precision} = \frac{TP}{TP + FP} \quad [1]$$

$$Recall = \frac{TP}{TP + FN} \quad [2]$$

A precision-recall curve was used to identify the optimal threshold that maximizes both metrics. This curve is a valuable gadget for enhancing the performance of a detection model. The metric of AP in Equation 3 is a summary of the precision-recall curve to indicate the mean of all precisions. The weighted mean of precisions at each threshold, which represents the increase in recall in comparison to the previous threshold, is used to calculate the AP.

$$\mathrm{AP} = \sum^{k=i=1} [R\,(\mathrm{k}) - R\,(\mathrm{k}+1)\,] \times P\,(\mathrm{k}) \qquad [3]$$

where i = number of thresholds, R = recall, k = index, and P = precision.

The mAP is formulated by averaging the AP across all classes, as shown in Equation 4.

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} APk \qquad [4]$$

where n = number of classes, k = index, and APk = the Average Precision of class k.

For this study, the mAP formulation is based on a single class: kitchen fires. There are two variations of metrics obtained from the base formula of mAP, also known as mAP@50 and mAP50-95.

However, these metrics also have limitations. Precision and recall do not account for the spatial accuracy of the bounding boxes, which can be critical in real-time fire detection scenarios. Additionally, mAP may not fully reflect the model's performance in safety-critical situations, where false negatives (missed fires) carry more severe consequences than false positives. While F1-score provides a better balance between precision and recall, it does not address how well the bounding boxes align with the actual fire, which can be crucial for accurate fire detection. Latency, or the time taken for the model to process an image and make a prediction, is another important metric for real-time applications, ensuring that the model can provide timely alerts in dynamic environments like kitchens. Despite these limitations, these metrics remain the standard in object detection research, providing a meaningful basis for comparative evaluation with existing models.

Dataset Acquisition

The kitchen fire data employed in this study are the original ones developed by the authors. Real video clips of kitchen fire scenarios were filmed in controlled conditions. Next, OpenCV was used to split the videos into still image frames of size 1920×1080 pixels and to filter out poor-quality or simply blurred frames. The ultimate dataset is in two versions,

one in RGB and another in HSV. The HSV dataset was built by converting each RGB image to HSV using OpenCV's cv2. cvtColor function. The dataset in this work contains images of visible kitchen fires, including flames on the stove and a pan under diverse lighting and background conditions. Thus, the focus of this work is on detecting flames during their visible state. Our experimental setup aims to study fire categorization in a kitchen setting, utilizing two different color spaces: RGB and HSV. A small, portable kitchen burner is used in the experiment as the primary source of fire, and data is collected from six different locations using two cameras: an iPhone X and a Samsung Galaxy Note 10 Lite. These smartphone models were deliberately chosen to reflect a realistic range of image capture devices available to everyday users. The iPhone X represents a high-end iOS device, while the Samsung Galaxy Note 10 Lite represents a mid-range Android device, allowing us to incorporate variations in hardware specifications, image processing algorithms, and camera sensors. This diversity helps enhance the generalizability and robustness of our vision-based fire detection model. These locations were carefully selected to ensure that our vision-based fire detection model would be robust, capturing thorough visual data from multiple perspectives.

The camera positioning is as follows: for the first round of recording, two cameras are placed at 45-degree angles relative to the stove's front view. Next is the placement of those two cameras, with one positioned directly in front and the other from a side view. These positions enable the capture of variations in fire dynamics and behavior, such as flame spread and intensity, that may occur during cooking activities. Furthermore, after the recording ended in the second round, the cameras were mounted at a top-down perspective, positioned 75 cm above the kitchen stove. This overhead view offers a crucial insight into the vertical spread of flames, which is vital for detecting fire patterns and understanding smoke evolution. The distance of the front camera from the stove is set at 50 cm, providing a clear, focused view of the fire's origin and progression.

The specific angles and camera arrangement for the experiment are shown in Figure 2. These camera placements are carefully configured to capture various angles and ensure comprehensive data collection, thereby improving the accuracy of fire detection systems. Due to the positioning of each camera, the experiment can examine flame properties such as color intensity and saturation, which are crucial for differentiating between fire and non-fire situations in actual kitchen settings. The experiment can achieve this by utilizing both RGB and HSV color spaces in combination.

By combining various viewpoints and color space analysis, this experimental design aims to minimize false positives while enhancing early fire detection and classification models in the kitchen context.

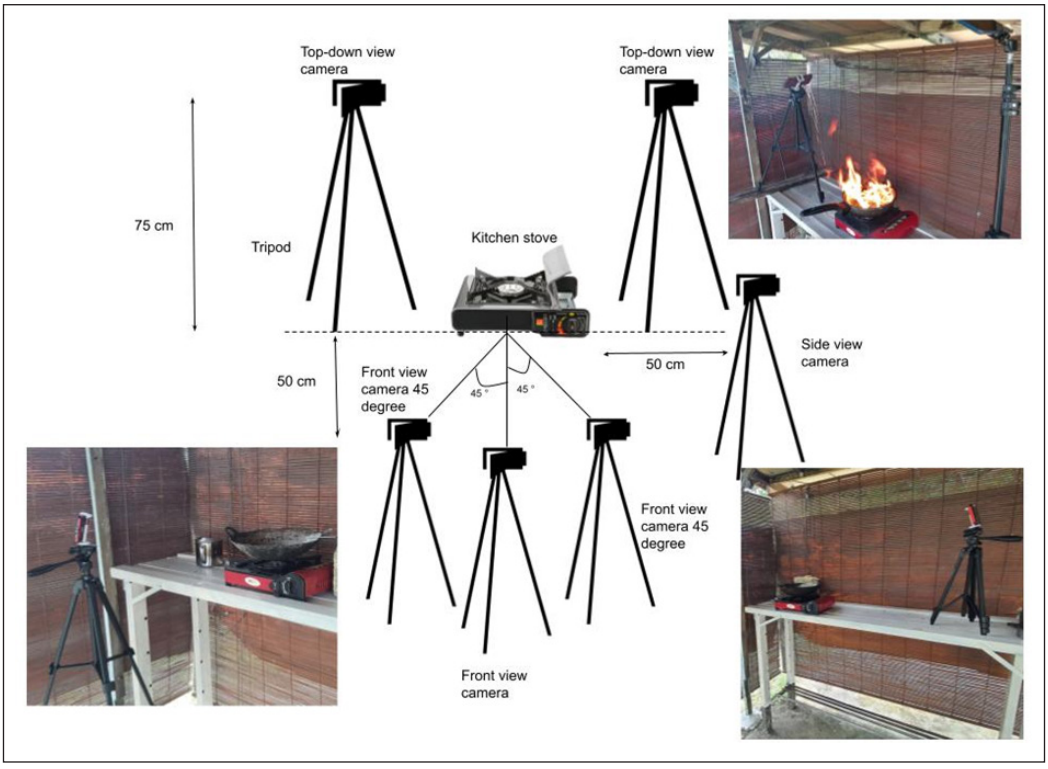


Figure 2. Camera positions in the experimental setup for fire detection

The images were taken from the six camera positions, as shown in Figures 3 and 4. In this experimental setup, the classification of two main types of fire: *stove flame* and *pan fire*. This classification is essential for differentiating between typical cooking flames, like those from a stove burner, and real fire incidents, like when a fire breaks out of a contained area and ignites objects like a pan. To improve the precision of fire detection in kitchen settings, the dataset produced by this experiment will concentrate on these two scenarios.

The term *stove flame* classification describes how a kitchen stove operates typically, with the flames contained within the burner and utilized for cooking. This class is crucial for determining safe cooking practices and lowering false alarms in fire detection systems. The *fire-in-the-pan* categorization, on the other hand, describes circumstances in which flames spread onto the pan, indicating a possible fire hazard. The identification of early fire occurrences that could result in uncontrolled kitchen fires depends heavily on this classification.

By capturing images from varying perspectives, from front, side, and top-down views, the dataset will provide comprehensive visual information for these two fire types. The combination of RGB and HSV color spaces will aid in identifying subtle differences in flame color, intensity, and spread between normal stove flames and hazardous pan fires.

This differentiation between *stove flame* and *pan fire* will play a vital role in refining the fire detection model, ensuring the system can accurately detect true fire hazards while minimizing false positives in kitchen environments.

Each position of the camera recorded a 10-minute video. The Google Colab platform's OpenCV package was then used to process these recordings. A total of 3,000 raw images were obtained by converting each video into a series of 500 images. Every image was thoroughly examined to guarantee the accuracy and caliber of the dataset. A final selection of 2,316 photos was made by eliminating blurry or poor-quality photos.

To evaluate the impact of color space on fire detection performance, the dataset was split into two versions: one in the original RGB color space and the other in HSV. The RGB images were used as-is for training the first model. For the HSV dataset, each RGB image was converted to HSV using OpenCV's cv2.cvtColor(image, cv2.COLOR_RGB2HSV) function in Python. The HSV-converted images were saved and labeled identically to the original RGB images to ensure consistency in the training process. Both the RGB and HSV image sets were then independently used to train the YOLOv5m model, using identical training parameters to facilitate a fair comparison. The YOLOv5 model was further tested on the selected dataset, with an emphasis on fire detection in both RGB and HSV color spaces.



Figure 3. A sample of six images represent six different positions of pan fire object from red, green, and blue (RGB) dataset: (a) 50 cm and 45° left, (b) 50 cm and 45° right, (c) 50 cm and side right, (d) 50 cm and center, (e) 75 cm and top left, and (f) 75 cm and top right, respectively

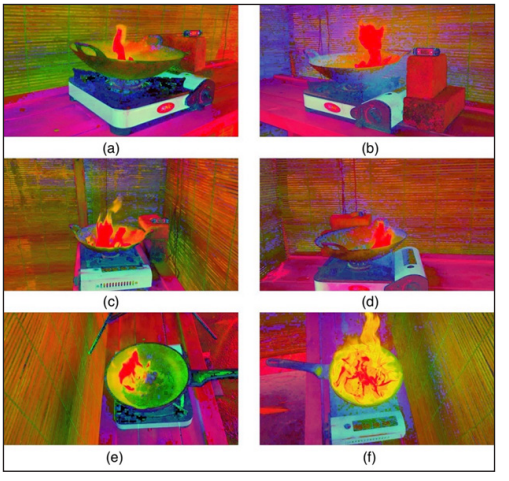


Figure 4. A sample of six images represent of six different positions of pan fire object from hue, saturation, and value (HSV) dataset: (a) 50 cm and 45° left, (b)50 cm and 45° right, (c) 50 cm and side right, (d) 50 cm and center, (e) 75 cm and top left, and (f) 75 cm and top right, respectively

Background Images

The background images with no fire indication from the stove and pan are shown in Figures 5 and 6. It has been suggested that including the background images in training might reduce false positives (Zhao et al., 2023). The six precise locations were used to get background images from the previous section.



Figure 5. A sample of six different background images from the red, green, and blue (RGB) dataset



Figure 6. A sample of six different background images from the hue, saturation, and value (HSV) dataset

Pre-Processing

The images were cropped to eliminate any unrelated backgrounds, specifically non-fire, as shown in Figure 7. The images were then resized after cropping to be compatible with

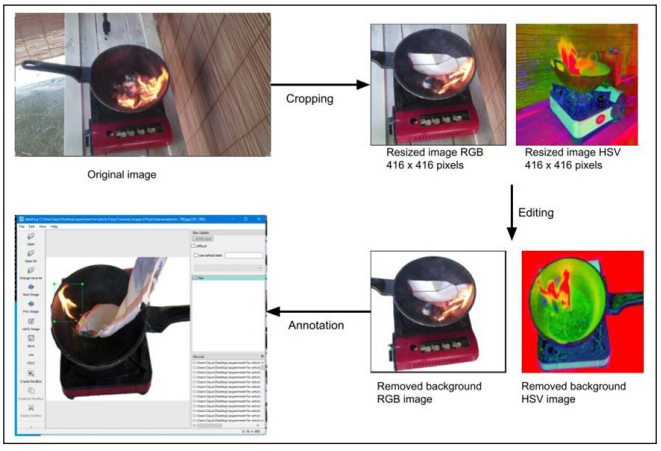


Figure 7. The flow of pre-processing steps for the fire image
Note. RGB = Red, green, and blue; HSV = Hue, saturation, and value

the YOLO model, which only accepts image sizes that are multiples of 32. The images were resized to a 416 × 416-pixel size.

In this study, no data augmentation techniques were applied during training. This decision was made to preserve the authenticity and raw visual characteristics of the kitchen fire scenarios as captured under real-world ambient lighting conditions. The aim was to evaluate the model's performance on naturally occurring variations in fire appearance without introducing synthetic transformations that could potentially distort fire features or misrepresent actual conditions. Additionally, no further normalization or pixel-intensity normalization was provided beyond the automatic internal pre-processing applied by the YOLOv5 framework to the input.

Dataset Partition

To ensure an exhaustive assessment of the model, the dataset was split into training, validation, and testing sets. The model was trained using the training set to identify the key features of the images. The validation set was used to monitor the model's performance throughout training, providing insights to tune its settings and parameters for improved performance. The performance of the trained model was assessed using the testing set. As Table 2 illustrates, the training, validation, and testing ratios in this article were set at 70:15:15. The datasets were shuffled using the scikit-learn shuffle function, with the random state parameter set to 1, to ensure reproducibility of the results. This ensures that the data is consistently and randomly ordered. There was no stratification splitting. However, randomization ensured that the two classes were equally distributed across the training, validation, and testing sets in the experiment. To investigate the effect of background context on model performance and the reduction of false positives, background images were integrated in Experiments 2 and 3 with a 5:1 fire-to-background ratio. Specifically, 244 background images were included in Experiment 2, and 486 background images were included in Experiment 3.

Table 2
Distribution of kitchen stove fire image dataset

Experiment	Color space	Number of images			Background images (5:1 ratio)
		Training	Validation	Testing	
1: Pan fire (No background)	RGB	200	43	43	0
	HSV	200	43	43	
2: Stove and pan fire with background	RGB	857	182	183	244
	HSV	857	182	183	244
3: Fused RGB and HSV	RGB and HSV	1,702	364	366	486

Note. RGB = Red, green, and blue; HSV = Hue, Hue, saturation, and value

Experiments

The experiments were conducted using a Lenovo ThinkStation P300 equipped with standard specifications, and all model training and testing were performed on Google Colab with Python 3.10.12 and PyTorch version 2.4.1+cu121. Training was conducted on an Nvidia (USA) Tesla T4 graphics processing unit (GPU) with 15 GB of video random access memory (VRAM). The central processing unit (CPU) and random-access memory (RAM) specifications were those provided by the Google Colab environment. The YOLOv5m model was selected for its balance between performance and computational efficiency. Required libraries included OpenCV, NumPy, PyTorch, and Matplotlib.

For YOLOv5m, it was trained for 200 epochs, using a batch size of 29, with the input image resolution of 416 × 416 pixels, which is similar to typical YOLO training settings. All training used the custom data set, and image preprocessing included resizing and normalization necessary for use with the YOLOv5 framework.

The model was optimized with the SGD optimizer with an initial learning rate of 0.01 and momentum of 0.937. An additional cosine learning rate schedule was applied, decaying the learning rate to 1% of its initial value at the end of training. Warmup was done with three epochs, initializing with a warmup momentum of 0.8 and a warmup bias learning rate of 0.1. These values correspond to the default YOLOv5 hyperparameters, which have been finely tuned for stable convergence.

The work from the first and second experiments is depicted in Figures 8 and 9. Figure 10 displays the block diagram for the third experiment. Before the training, validation, and testing stages, pre-processing was applied to every stove fire and pan fire, including the background images. To start, every fire on the pan was individually cropped to remove any non-fire items. In terms of the background images, any fire objects that appeared in them were removed by cropping off the area of the image showing the pan and stove on fire, resulting in a clean background. The bounding box surrounding the targeted items was then provided as ground truth using LabelImg to annotate the fire images. The background images were not labelled and were simply added to the training dataset.

In our approach, the original YOLOv5m model provided by Ultralytics was used without any architectural changes. The model was initialized with pre-trained common objects in context (COCO) weights to take advantage of transfer learning gains. The structure of the backbone, neck, and detection head was not modified based on the model. The standard version has been used with the intention that, in case of fire occurrence, it would be a fair comparison for the RGB and HSV scenarios.

The datasets were then divided into three sets of experiments:

1. In the first experiment, the study concentrated solely on pan fire images without any kitchen background. The purpose of this experiment was to determine whether color space (RGB vs. HSV) had an impact on fire detection accuracy. The RGB and HSV

datasets were trained separately on YOLOv5m, and the results were compared to determine the color space with superior performance in a controlled, background-free environment.

2. The second set of experiments consists of both stove fire and pan fire images. This experiment compares the performance of RGB and HSV on background images. The performance of the RGB and HSV datasets, which included photos of stoves and pan fires, was examined after each dataset was trained separately.

3. The third experiment uses the fusion of both the RGB and HSV datasets to create more training sets. This experiment aims to investigate whether the model performs better at fire detection when it learns features from both RGB and HSV datasets simultaneously.

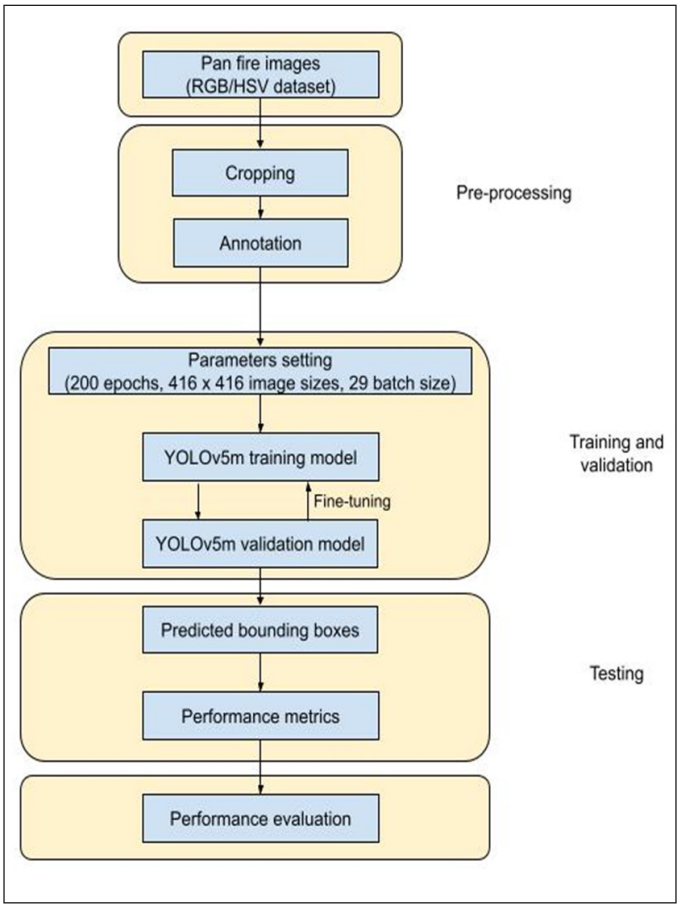


Figure 8. Flowchart of the first experiment
Note. RGB = Red, green, and blue; HSV = Hue, saturation, and value; YOLOv5m = You Only Look Once version 5 medium

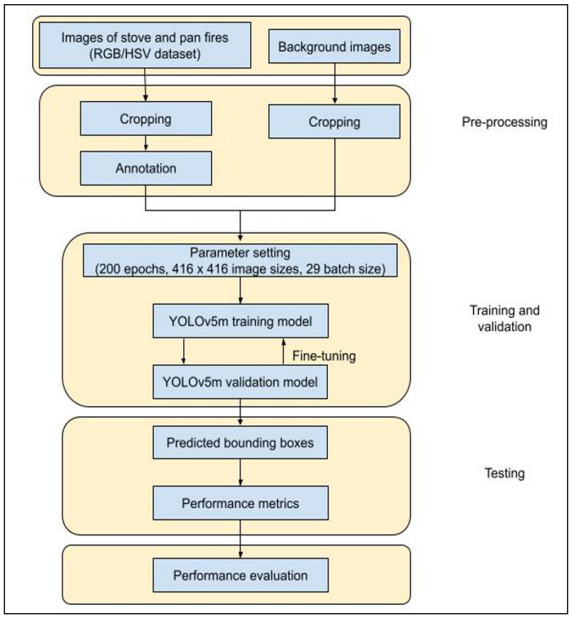


Figure 9. Flowchart of the second experiment
Note. RGB = Red, green, and blue; HSV = Hue, saturation, and value; YOLOv5m = You Only Look Once version 5 medium

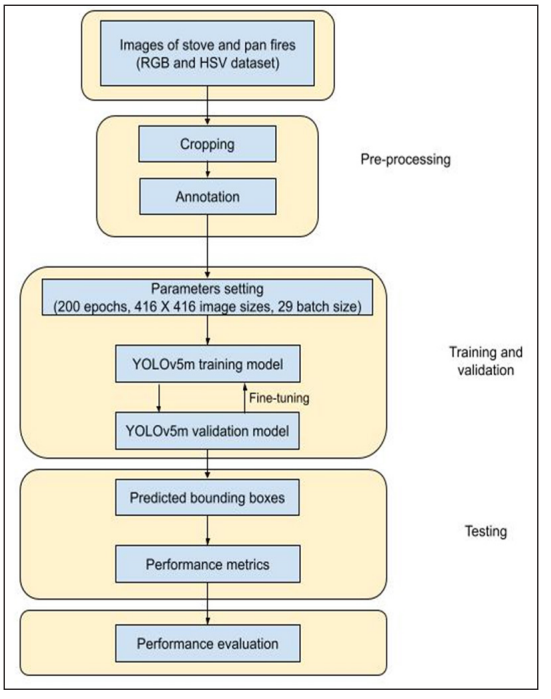


Figure 10. Flowchart of the third experiment
Note. RGB = Red, green, and blue; HSV = Hue, saturation, and value; YOLOv5m = You Only Look Once version 5 medium

RESULTS AND DISCUSSION

Results Overview

The performance of the RGB, HSV, and fused RGB-HSV datasets is compared across important assessment measures, including recall, precision, F1-score, mAP@50, and mAP@50-95, in the Table 3, which summarizes the experiment results. These metrics provide information about how well the model detects different scenarios and an overall sense of how effectively each model can detect fire and stove-related events across different kitchen settings. The calculations of the metrics were based on the validation functions provided in the YOLOv5 platform. A prediction was a true positive if the IoU of the predicted and ground truth bounding box was 0.5 or larger. mAP@50, which is the average precision at an IoU threshold of 0.5, and mAP@50–95 reflects the robustness of the model by averaging the performance over the IoU ranging from 0.5 to 0.95 with an increment of 0.05.

Table 3
Overall result of performance metrics for three experiments

Experiment	Color space	Recall	Precision	F1-score	mAP@50	mAP@50-95
Experiment 1 (Pan fire)	RGB	0.843	0.886	0.863	0.873	0.651
	HSV	0.882	0.931	0.905	0.938	0.632
Experiment 2 (Stove and pan fires)	RGB	0.933	0.941	0.937	0.950	0.632
	HSV	0.921	0.966	0.943	0.953	0.624
Experiment 3 (Fusion RGB-HSV)	Fused RGB and HSV	0.918	0.953	0.935	0.949	0.622

Note. mAP = Mean average precision; RGB = Red, green, and blue; HSV = Hue, saturation, and value

Experiment 1: Training Dataset without Background Images

An extensive analysis of our suggested model's performance on datasets that only contain specific fires on pan images in Table 4 . Various performance metrics were used to analyze the capabilities of the model and provide an extensive discussion of the results.

Table 4
Performance metrics for Experiment 1

Training dataset	Recall	Precision	F1-score	mAP@50	mAP@50-95
RGB without background images	0.843	0.886	0.863	0.873	0.651
HSV without background images	0.882	0.931	0.905	0.938	0.632

Note. mAP = Mean average precision; RGB = Red, green, and blue; HSV = Hue, saturation, and value

In the first experiment, fire detection accuracy was tested in a controlled background-free environment by comparing RGB and HSV color spaces. The analysis was conducted based on the performance of YOLOv5m, which was trained separately on RGB and HSV datasets of pan fire images. The findings provide insight into how the two-color spaces differ in terms of processing and performance when it comes to detecting kitchen fires.

Processing Time and Epochs

For 200 epochs, or an equal number of iterations over the datasets, the RGB and HSV models were trained. Both color spaces required equal computational resources for training, as indicated by the consistency of processing times, which were approximately 15 minutes (0.249 hours for RGB and 0.248 hours for HSV). This equivalency shows that the amount of time needed to train the YOLOv5m model in this controlled setting was not significantly affected by the color space selection.

Inference Speed

Between the two models, there was a small difference in the inference speed during validation. While the HSV dataset processed photos at 2.32 iterations per second (it/s), the RGB dataset processed images at 2.41 it/s. Despite being negligible, this difference implies that the HSV photos were validated a little bit faster. This might be explained by the way the YOLOv5m model handles the color channels of RGB and HSV differently, albeit the differences are not great enough to imply that one has a computational advantage over the other.

Performance Metrics

The main objective of this experiment was to evaluate the accuracy of fire detection between the RGB and HSV color spaces. The performance metrics in Table 4 shows a considerable difference between the HSV and RGB models. In almost every metric used in this experiment, the HSV color space outperformed RGB. Compared to RGB, which obtained a recall of 0.843, a precision of 0.886, an F1-score of 0.863, and an mAP@50 of 0.873, HSV obtained a higher recall of 0.882, a precision of 0.931, an F1-score of 0.905, and an mAP@50 of 0.938. This shows that removing the background improves HSV's ability to identify fire from non-fire elements. This might be because the nature of the HSV color space allows less interference in color feature separation, such as the hue of fire in the presence of minimal distractions in the background. Nonetheless, it is intriguing to observe that, although the mAP@50-95 (which is a more rigorous measure and assessed over a range of IoUs) is a little lower for HSV (0.632) than RGB (0.651), the precision and recall metrics still favor HSV. This suggests that even though RGB may be constructed with greater leeway to varying levels of IoU, its effectiveness is inferior when it comes to detecting fire, in which case it is restricted to a narrow range of thresholds, whereby it facilitates detection more accurately than the RGB model.

Nonetheless, although the mAP@50-95, which is a more rigorous measure that assesses over a range of IoUs, is slightly lower for HSV (0.632) than for RGB (0.651), the precision, recall, and F1-score metrics still favor HSV. That means that HSV is more accurate in identifying fire when restricted to certain thresholds, even though RGB might have a wider tolerance at different IoU values. As demonstrated in Figures 11 and 12, the YOLO5m is utilized to detect the validation and testing image samples, respectively.

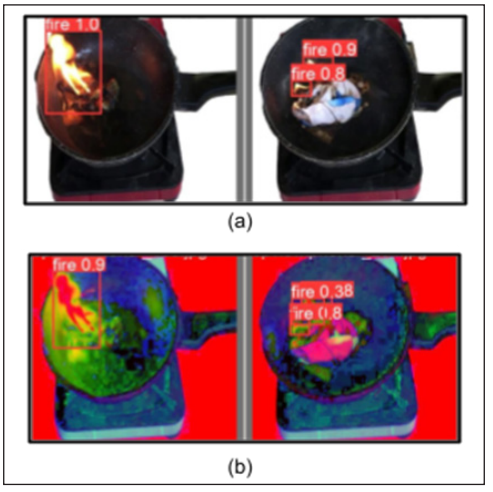


Figure 11. The YOLOv5m model was used to detect validation samples of (a) RGB and (b) HSV images, respectively
Note. YOLOv5m = You Only Look Once version 5 medium; RGB = Red, green, and blue; HSV = Hue, saturation, and value

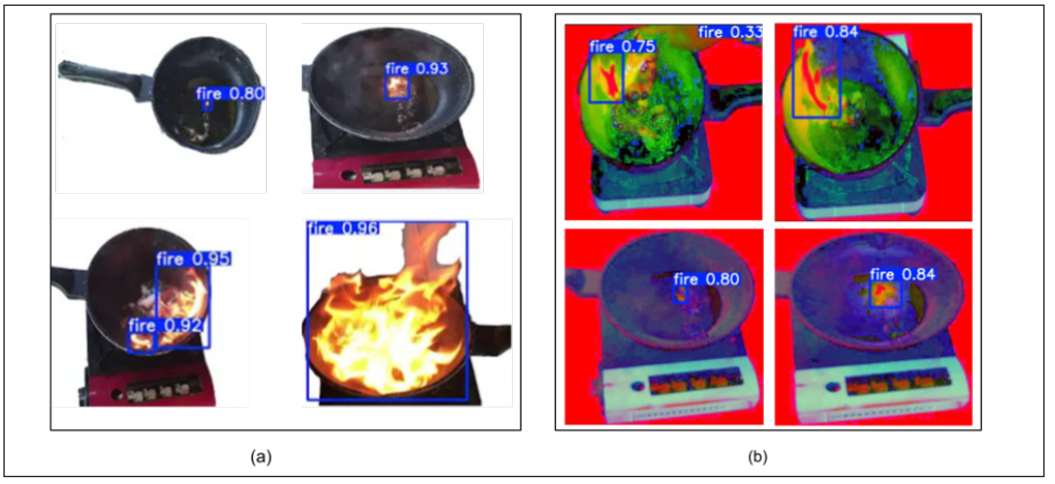


Figure 12. The YOLOv5m model was used to detect testing samples of (a) RGB and (b) HSV images, respectively
Note. YOLOv5m = You Only Look Once version 5 medium; RGB = Red, green, and blue; HSV = Hue, saturation, and value

Experiment 2: Training Dataset with Background Images

The second experiment was conducted with an expanded scope of the dataset by incorporating both stove flame and pan fire images, including the background surrounding, which increased the task complexity. The objective was to determine the performance matrix differences between RGB and HSV in the first experiment with more complex, real-world scenarios involving kitchen backgrounds. Both RGB and HSV color spaces were evaluated using the YOLOv5m model based on the detection of fire and stove elements, as shown in Table 5.

Table 5
Performance metrics for Experiment 2

Training dataset	Recall	Precision	F1-score	mAP@50	mAP@50-95
RGB with background images	0.933	0.941	0.937	0.950	0.632
HSV with background images	0.921	0.966	0.943	0.953	0.624

Note. mAP = Mean average precision; RGB = Red, green, and blue; HSV = Hue, saturation, and value

Processing Time and Epochs

Both the RGB and HSV datasets required more time to train than in the first experiment, with RGB taking about 46 minutes (0.771 hours) and HSV taking a longer duration at 49 minutes (0.808 hours). Training was done for both datasets over 200 epochs. Given the complexity of the dataset and the inclusion of background features, the longer training time was anticipated. Given that the same number of epochs was used to train both models, the significantly longer processing time for HSV implies that the background complexity may have contributed to a slightly slower training process for this color space.

Inference Speed

Comparing RGB (1.28 it/s) to the first experiment, the inference speed during validation was noticeably slower. This is probably because there were more different kinds of images, including backgrounds, stove, and pan fire, which raised the computational processing demand. HSV model performance, on the other hand, was consistent with the first experiment, demonstrating an inference speed of 2.38 it/s, nearly twice as fast as the RGB model. This suggests that the HSV color space can handle backdrop clutter with lower processing time.

Performance Metrics

Both RGB and HSV color spaces performed well overall in terms of fire detection abilities. Nevertheless, a review of the performance measures revealed some significant variations. The HSV model outperformed RGB with a precision of 0.966 as opposed to

0.941. This indicates that the HSV model is more accurate in differentiating between fire and stove elements from the background, as it generates fewer false positives. For real-world applications, where lowering false positives is essential, this increased precision is noteworthy. With a recall score of 0.933 against HSV 0.921, the RGB model scored somewhat better than HSV. This indicates that, when compared to HSV, the RGB model was more accurate in recognizing true positives, catching a higher proportion of real fire and stove occurrences, but at the cost of more false positives. In terms of mAP@50, RGB scored 0.950 and HSV achieved 0.953, indicating equivalent performance for both models. According to this metric, the two-color spaces identify fire occurrences at the 50% IoU threshold nearly equally well. Nonetheless, the slightly elevated score of the HSV model bolsters its advantage in precision. In the mAP@50-95 metric, the RGB achieved a higher score of 0.632 than the HSV model with 0.624. This finding suggests that RGB may perform better across a range of IoU thresholds, potentially enhancing its robustness in identifying fires in congested environments with diverse background objects and surroundings.

Class-Wise Breakdown

More information about how each color space handles pan fire, and stove images can be found in the performance metrics broken down by class, shown in Table 6.

Table 6
Performance metrics for Experiment 2 by class

Model	Training class	Recall	Precision	F1-score	mAP@50	mAP@50-95
RGB	Pan fire	0.887	0.876	0.881	0.909	0.567
	Stove flame	0.994	0.989	0.992	0.991	0.697
HSV	Pan fire	0.939	0.852	0.893	0.919	0.558
	Stove flame	0.993	0.989	0.991	0.988	0.691

Note. mAP = Mean average precision; RGB = Red, green, and blue; HSV = Hue, saturation, and value

For pan fire images: The HSV model performed better than the RGB model (0.887) in terms of recall (0.939), indicating that it could identify more actual fires. HSV achieved mAP@50-95 (0.558) and precision (0.852), which were, however, lower than those of RGB (0.876, 0.567 for mAP@50-95). As a result, HSV provided a greater range of detection performance, even though it was more consistent in spotting fire. It also created more false positives. On the other hand, the RGB model was able to achieve a better balance between precision and recall metrics, because it was more adept at differentiating colors over complex backgrounds.

Stove flame images: Both color spaces performed well. While the HSV model produced similarly good metrics (recall = 0.993, precision = 0.989), the RGB model obtained a recall of 0.994 and a precision of 0.989. The nearly flawless detection of stoves in all models

suggests that significant visual cues, including the stove's unique shape and structure, were crucial in achieving high detection accuracy across all color spaces.

Both classification detection utilizing YOLOv5m for both color spaces are shown in Figures 13 and 14.

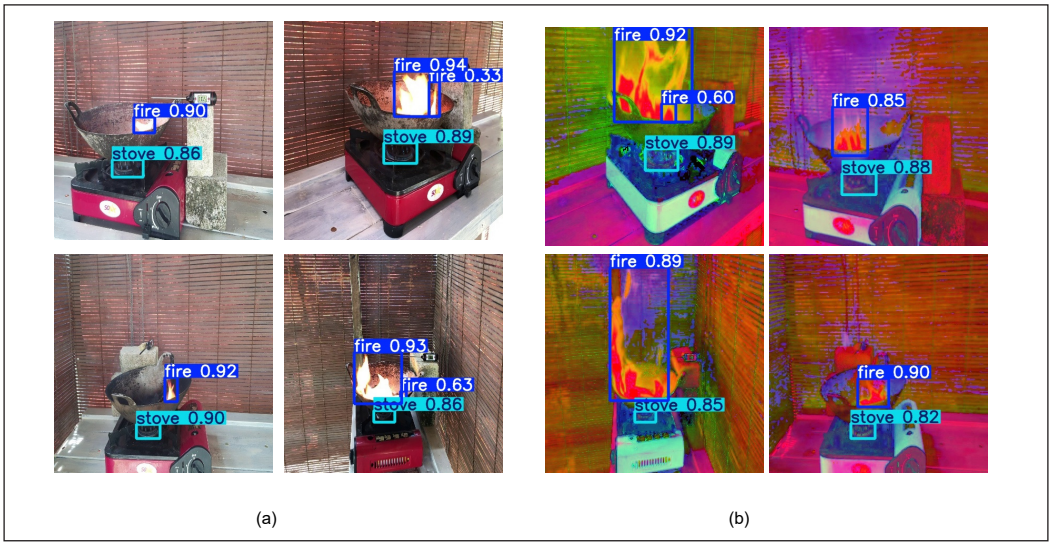


Figure 13. The YOLOv5m model was used to detect validation samples of (a) RGB and (b) HSV images, respectively

Note. YOLOv5m = You Only Look Once version 5 medium; RGB = Red, green, and blue; HSV = Hue, saturation, and value

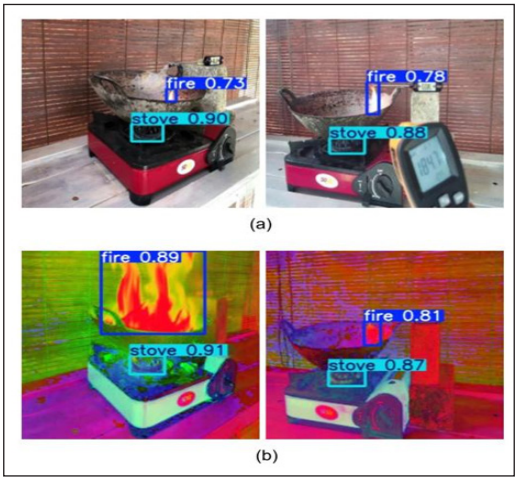


Figure 14. The YOLOv5m model was used to detect testing samples of (a) RGB and (b) HSV images, respectively

Note. YOLOv5m = You Only Look Once version 5 medium; RGB = Red, green, and blue; HSV = Hue, saturation, and value

Experiment 3: Training Dataset Fused RGB-HSV with Background Images

The third experiment was done to evaluate the performance of a fusion of two different color spaces in one dataset. Through the combination of both RGB and HSV datasets, the experiment aimed to take advantage of RGB and HSV complementing capabilities, particularly in handling background clutter and color differentiation.

Training Process and Epochs

In contrast to the 200 epochs used in earlier experiments, the model was trained for only 153 epochs. The training period was completed in about 69 minutes, or 1.15 hours, with an early halt happening when no improvement was seen for the final 10 epochs. This early stopping mechanism indicates that the model reached its optimal solution more quickly. However, it also suggests that longer training times were not necessary to achieve peak performance, due to the added complexity introduced by combining the RGB and HSV datasets.

Inference Speed

During validation, the inference speed was 1.19 iterations per second (it/s), which was slower than in Experiment 2 (2.38 it/s) for HSV. The larger and more complex dataset produced by combining two color spaces is likely the cause of this speed reduction. More computational overheads may have been incurred by the model's need to process a more varied set of features.

Performance Metrics

The fusion model did not significantly exceed the RGB or HSV models individually, but it did maintain good performance, especially in precision. The following is a summary of key performance metrics as shown in Table 7.

Table 7
Performance metrics for Experiment 3

Training dataset	Recall	Precision	F1-score	mAP@50	mAP@50-95
Fusion of RGB and HSV datasets	0.918	0.953	0.935	0.949	0.622

Note. mAP = Mean average precision; RGB = Red, green, and blue; HSV = Hue, saturation, and value

As in the earlier experiments, a precision of 0.953 shows that the fusion model maintained its high level of accuracy in lowering false positives. This great precision indicates that the fusion did not negatively impact the model's capacity to accurately identify stove or real fire incidents, as it is consistent with the individual RGB and HSV models.

The recall value of 0.918 shown in this experiment is marginally less than in Experiment 2 (HSV = 0.921, RGB = 0.933). This implies that while the fusion model performed well, it was not as successful as the individual RGB and HSV models in detecting every fire occurrence. The enhanced complexity of the fused dataset may have added more variability and confusion in identifying specific fire patterns, which could account for the slightly lower recall.

The mAP@50 result with 0.949 shows a good overall detection capacity, nearly matching the results from the individual RGB (0.950) and HSV (0.953) models. This consistency indicates that, at the 50% IoU threshold, combining RGB and HSV does not impair the model's performance, suggesting that it remains capable of detecting fire and stove elements.

The mAP@50-95 value of 0.622 is marginally less than the mAP@50-95 values obtained from the separate RGB (0.632) and HSV (0.624) experiments. This indicates that the fusion model struggled at higher IoU thresholds, where bounding box predictions require greater accuracy. This could imply that some inconsistencies or overlapping features were introduced when merging features from the two-color spaces, which would have resulted in less accurate predictions in difficult circumstances.

Class-Wise Breakdown

Important details on how the fusion model processed fire and stove images may be found in the class-by-class breakdown shown in Table 8.

Table 8
Performance metrics for Experiment 3 by class

Model	Training class	Recall	Precision	F1-score	mAP@50	mAP@50-95
Fusion of RGB and HSV datasets	Pan fire	0.913	0.847	0.879	0.906	0.561
	Stove flame	0.993	0.989	0.991	0.991	0.684

Note. mAP = Mean average precision; RGB = Red, green, and blue; HSV = Hue, saturation, and value

Pan fire images: In contrast to the RGB (0.887) and HSV (0.939) models from the second experiment, the recall for pan fire images was, on average, 0.913. This implies that the fusion model's ability to identify actual fire incidents was effective. The fusion model, however, generated more false positives, as evidenced by the decreased precision for fire images (0.847) compared to the RGB model (0.876). Furthermore, the mAP@50-95 for fire images (0.561) was on average of RGB (0.567) and HSV (0.558), suggesting a deterioration in the model's capacity to locate fire incidents at various IoU thresholds reliably.

Stove flame images: The recall and precision remained extremely high (recall = 0.993, precision = 0.989), in line with other trials. The fusion of RGB and HSV did not adversely

affect the accuracy of stove detection, as evidenced by the mAP@50-95 for stove images of 0.684, which was slightly lower than the prior results in Experiment 2. This implies that the complexity added by the fused dataset had a lesser impact on stove flame photos, which are easier to distinguish and identify visually.

Both classification detection utilizing YOLOv5m for experiment 3 is shown in Figures 15 and 16 during the validation and testing process.

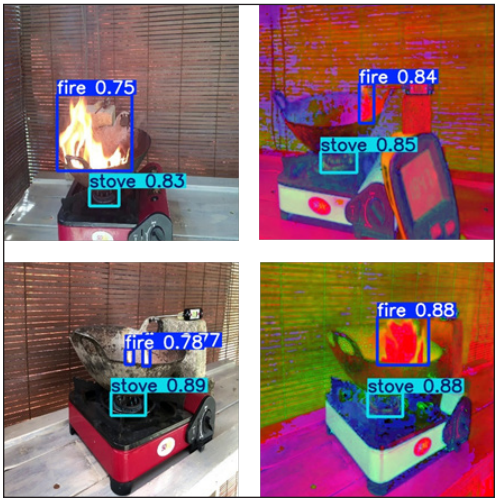


Figure 15. The YOLOv5m model was used to detect validation samples of fused RGB and HSV images, respectively
Note. YOLOv5m = You Only Look Once version 5 medium; RGB = Red, green, and blue; HSV = Hue, saturation, and value

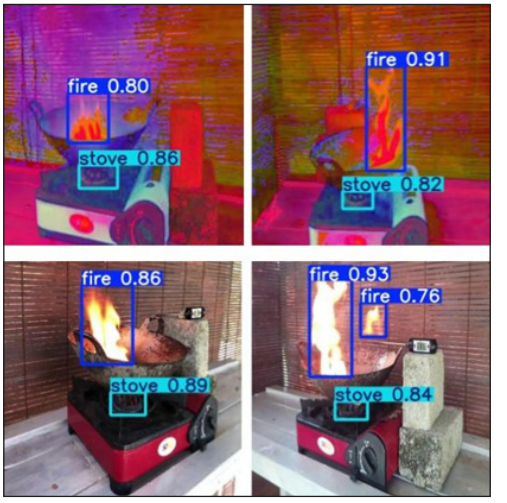


Figure 16. The YOLOv5m model was used to detect testing samples of fused RGB and HSV images
Note. YOLOv5m = You Only Look Once version 5 medium; RGB = Red, green, and blue; HSV = Hue, saturation, and value

Analysis of Confusion Matrices for Each Experiment

Confusion matrices are used to further evaluate model performance and understand the nature of classification errors in these experiments. These matrices demonstrate the model's ability to distinguish between 'fire', 'stove fire', and 'background' classes. A higher incidence of false positives, specifically, misclassifications of background as fire, is shown in Experiment 1 (false positive rate: 22% for HSV, 23% for RGB), where background images were excluded. In contrast, Experiments 2 and 3, which incorporated varied background contexts, show a marked reduction in false positives, underscoring the critical role of background information in improving model robustness and reducing misclassification rates.

In the first experiment, which involved no background, models were trained on images depicting pan fire, allowing isolated evaluation of the impact of color space on fire detection. The confusion matrices for RGB and HSV in Figure 17 show strong diagonal dominance,

indicating that the models were highly accurate in classifying fire versus background. The HSV database model yielded fewer false negatives than the RGB model. This suggests that HSV color space provides superior color segmentation properties during the detection of fire pixels when minimal noise is involved.

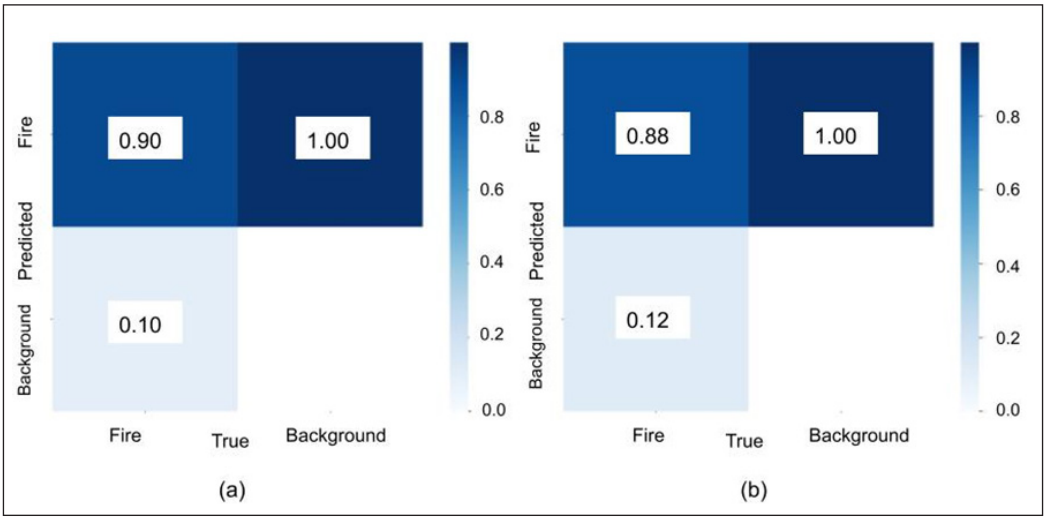


Figure 17. Confusion matrices for the first experiment using (a) RGB and (b) HSV images, respectively
Note. RGB = Red, green, and blue; HSV = Hue, saturation, and value

In the second experiment, the dataset included a mix of stove fires and pan fires, as well as actual kitchen environments. This is the challenging part of the model as it involves a complex background. As shown in Figure 18, both HSV and RGB models show increased misclassification rates. The increment of misclassified stove fire images as background in the RGB model suggests that its features were more susceptible to distraction from surrounding kitchen elements. However, the HSV model depicted improvement in these scenarios, maintaining clearer separation between fire-related classes and the background. The HSV color space shows better color contrast and stability under varying lighting conditions, which is normal in a real kitchen environment.

The third experiment was executed to assess the fused RGB and HSV dataset performance, which consisted of training samples drawn from both RGB and HSV formats. The result from the confusion matrix in Figure 19 demonstrates that the model trained outperformed individual color space models. A few pan fire and stove fire images were misclassified together with background images. This finding confirms that feature diversity obtained through fusion allows the model to learn a more comprehensive representation of fire characteristics accurately.

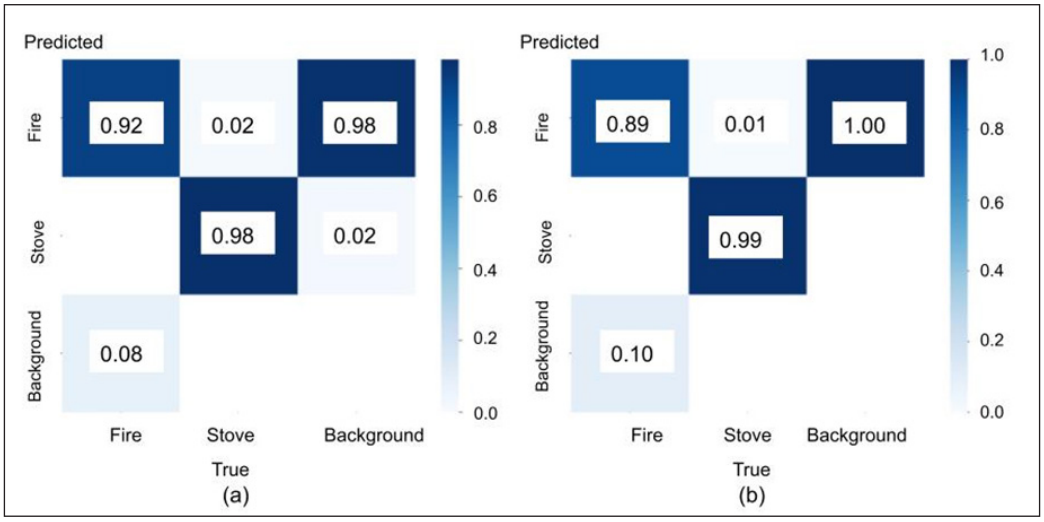


Figure 18. Confusion matrices for the second experiment using (a) RGB and (b) HSV images, respectively
Note. RGB = Red, green, and blue; HSV = Hue, saturation, and value



Figure 19. Confusion matrix for the third experiment using fused RGB and HSV images
Note. RGB = Red, green, and blue; HSV = Hue, saturation, and value

The results revealed that HSV was slightly better than RGB in some cases. This effect might also be related to the fact that HSV allows the direct separation of chromatic (hue and saturation) and intensity (value) information and is thus more resilient to lighting variations. This property likely helps HSV better localize fire features in kitchen scenes, where illumination changes are common due to flames, reflections, or shadows.

Thus, the preference for HSV over RGB found in Experiments 1 and 2 might be because it separates chromatic content (hue) from illuminant (value). Fire is usually very red, orange, and yellow, which is consistent in all lighting. Unlike RGB, where color and brightness are intertwined, HSV helps the fire model to concentrate more on the hue feature of the fire in distinguishing flames from typical kitchen backgrounds, such as metal surfaces, tiles, and utensils.

Furthermore, with background images in Experiment 2, the task became more complex, possibly including items such as stove knobs, cooking utensils, or reflections of lighting. The latter might introduce ambiguities to the model in RGB space, as similar colors can also exist in non-fire areas. Unlike HSV, the color information separation on the chromatic side of the HSV model that was applied in the framework subsequently suppressed the background clutter and made detection more robust.

The false positive rate, which is defined as the proportion of images without fire that were misclassified as containing fire, was significantly higher in Experiment 1 (0.22 for HSV, 0.23 for RGB), with reduced background variability. However, the false positive rate dropped dramatically in Experiments 2 and 3, using a variety of background images (0.13 for HSV, 0.08 for RGB, and 0.08 for fused RGB + HSV). These results validate that the use of background images enhances the model's ability to discriminate between fire and non-fire scenes, resulting in a decrease in the number of false alarms.

Comparison Between the Proposed Model and Other Approaches

A comparison between the proposed YOLOv5m model and other cutting-edge YOLOv5 techniques used in current fire detection studies is shown in Table 9. All the studies that were utilized for comparison involved RGB datasets, while one study used HSV and luminance, chrominance-blue, and chrominance-red (YCbCr) datasets as part of their training images. In general, the comparison illustrates the potential advantages of utilizing both RGB and HSV data for object detection tasks, particularly for fire detection. The results shown in Table 9, the greatest mAP@0.5 values, 0.950 and 0.953, respectively, are obtained by the proposed model YOLOv5m with background images (HSV and RGB), suggesting good precision for detecting fire events with background context. The performance of the proposed models demonstrates a trade-off between the use of background and non-background images. It is interesting to note that RGB and HSV with background images maintain greater mAP@0.5 but have difficulty with mAP@0.5:0.95, suggesting a tendency to over-rely on background information or difficulties generalizing across different IoUs. HSV-based models typically outperform RGB-based models or perform about the same. For instance, compared to the RGB equivalent (0.873 mAP@50), the suggested model with HSV, excluding the background (0.938 mAP@50), performs better. This implies that HSV, perhaps because of its more intuitive representation of color fluctuations in fire, might be more suitable for fire detection.

Table 9
Comparison between the proposed model and other approaches

Method	Color space	mAP@050	mAP@50-95	Type of fire detection
YOLOv5 and flame threshold segmentation (Zhao et al., 2022)	RGB	0.863	0.576	Fire, smoke, lamp, reflections, sunshine
YOLOv5s-CA (W. Liu et al., 2022)	RGB	0.946	-	Smoke and fire
YOLOv5 with SE attention mechanism (Fan & Zhan, 2023)	RGB	0.941	0.799	Flame
YOLOv5 and YOLOv8 (Mahmoud et al., 2024)	HSV and YCbCr	0.895	0.954	Fire and smoke
YOLOv5m (Proposed model)	RGB without background images	0.873	0.651	Fire
YOLOv5m (Proposed model)	HSV without background	0.938	0.632	Fire
YOLOv5m (Proposed model)	RGB with background images	0.950	0.632	Fire
YOLOv5m (Proposed model)	HSV with background images	0.953	0.624	Fire
YOLOv5m (Proposed model)	Fusion of RGB and HSV datasets	0.949	0.622	Fire

Note. mAP = Mean average precision; YOLOv5 = You Look Only Once version 5 ; RGB = Red, green, and blue; YOLOv5s-CA = You Look Only Once version 5 small with coordinate attention; SE = Squeeze-and-excitation; YOLOv8 = You Look Only Once version 8; YCbCr = Luminance, chrominance-blue, and chrominance-red; YOLOv5m = You Look Only Once version 5 medium; HSV = Hue, saturation, and value

Most models, including the proposed one, exhibit a rapid decline in mAP@0.5 to mAP@0.5:0.95, highlighting areas that require improvement for handling diverse object scales and ensuring reliable detection accuracy. This disparity implies that although the models are capable of simple fire detection, they are not as good at more complex detections across a range of IoU thresholds. The small benefits of combining RGB and HSV color spaces imply that the complementary qualities of these color spaces are not being fully utilized. Subsequent research could examine sophisticated fusion methodologies or examine the reasons behind the fusion's inability to produce appreciably superior outcomes.

Implications and Ethical Considerations

This work demonstrates that using the color space method along with YOLOv5 will lead to improved kitchen fire detection, thus promoting household safety by providing timely responses in fire emergencies. The method is applicable for low-cost, real-time use in

the home, where conventional systems fail. Legally, the issue of privacy is paramount, especially when it comes to using cameras in private locations. Local data analysis and user consent are critical for protecting personal privacy and complying with data protection laws.

CONCLUSION

The performance of object detection models can be greatly influenced by the selection of the training dataset. In particular, the top-performing dataset changes based on the metric of interest. Combining datasets can improve overall performance, demonstrating the advantages of using diverse training data sources in object detection applications. To determine the best training datasets and methods for object detection models, these results highlight the significance of carrying out experiments and evaluations. However, in this study, the proposed model resulted in the best overall mean performance in the first experiment, RGB, without background images, generating the highest mAP@0.5:0.95 score of 0.651. Interestingly, the overall performance in terms of precision and recall still favors HSV, even while the mAP50-95 is lower when compared to RGB. This suggests that HSV is more accurate in detecting fire when restricted to thresholds, whereas RGB may have a wider tolerance at different IoU values. Nevertheless, the fused model demonstrated high performance in identifying stove images (recall = 0.993, precision = 0.989), suggesting that stove identification is less susceptible to changes in color space and can be consistently achieved using RGB, HSV, or both. The detection scenario determines which RGB and HSV to use. HSV gives better precision for more straightforward, controlled conditions, which is essential for reducing false positives. However, RGB can more clearly distinguish fire from surrounding objects in real-world kitchen settings with more complicated backgrounds, and it might offer more dependable detection in certain situations. To analyze the benefits of combining a color space, we adopted an early fusion strategy. More specifically, each image was transformed from RGB to HSV, and both (RGB and HSV) were stacked along the channel dimension to form a single multichannel input. The YOLOv5 model was then trained using this aggregated input. The aim was to get the best attributes of each color spectrum to describe fire objects. However, the results show that this embedding was not beneficial, possibly because the input is more complex.

The proposed datasets were assessed using the specific YOLOv5m architecture model. In addition to evaluating the dataset using other state-of-the-art models, apart from YOLOv5, future research should involve increasing the quantity and variety of images and testing the model's robustness under various settings to further improve its performance. Future research could further examine the robustness of the proposed model at various distances from the camera to distinguish between fire and other reddish objects. This study was limited to a single run of each experiment due to time and computational constraints. As a result, statistical measures such as mean, standard deviation, and significance testing

between the RGB and HSV results were not performed. Future work will involve repeated experiments to facilitate more robust statistical analysis. Another limitation is that the study focused on controlled kitchen environments. Further research is needed to evaluate the performance of the proposed approach in more complex and cluttered real-world kitchens. The dataset used in this study, while carefully curated, may not fully represent the wide variety of fire scenarios that can occur in kitchens. Future work should explore the use of larger and more diverse datasets.

The research also suggested employing more advanced GPU specifications to reduce training times and investigating how various image resolutions impact the model's performance. To further improve object detection model performance, additional feature engineering can be investigated, in line with the indicated future work. Another suggestion is to investigate performance in real-world kitchens, exploring the use of temporal information, evaluating computational efficiency on embedded devices, and addressing the limitations of the current dataset. By offering more details about an object's form and texture, these characteristics can raise the precision of object detection algorithms. To further confirm the durability of HSV in practical applications, additional research may examine how these results apply to more complex scenarios with background noise.

ACKNOWLEDGEMENTS

The authors sincerely thank the Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, for supporting this project.

REFERENCES

Abdusalomov, A. B., Mukhiddinov, M., Kutlimuratov, A., & Whangbo, T. K. (2022). Improved real-time fire warning system based on advanced technologies for visually impaired people. *Sensors*, *22*(19), 7305. https://doi.org/10.3390/s22197305

Bakri, N. S., Adnan, R., Samad, A. M., & Ruslan, F. A. (2018). A methodology for fire detection using colour pixel classification. In *IEEE 14th International Colloquium on Signal Processing and Its Applications* (pp. 94–98). IEEE. https://doi.org/10.1109/CSPA.2018.8368692

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). *YOLOv4: Optimal speed and accuracy of object detection*. arXiv. http://arxiv.org/abs/2004.10934

Bu, F., & Gharajeh, M. S. (2019). Intelligent and vision-based fire detection systems: A survey. *Image and Vision Computing*, *91*, 103803. https://doi.org/10.1016/j.imavis.2019.08.007

Choi, H.-B., Hwang, E.-H., & Choi, D.-M. (2023). Indoor air quality sensor utilization for unwanted fire alarm improvement in studio-type apartments. *Fire*, *6*(7), 255. https://doi.org/10.3390/fire6070255

Chow, W. K., & Liu, J. (2016). Application of nonlinear dynamics in studying flashover fire in a small open kitchen. *Journal of Applied Mathematics and Physics*, *4*(5), 914–924. https://doi.org/10.4236/jamp.2016.45100

Dilshad, N., Khan, T., & Song, J. (2023). Efficient deep learning framework for fire detection in complex surveillance environment. *Computer Systems Science and Engineering*, *46*(1), 749–764. https://doi.org/10.32604/csse.2023.034475

Fan, J., & Zhan, X. (2023). Flame recognition algorithm for aerial cruising firefighting robot based on RGB and improved YOLOv5. In *4th International Conference on Computer Vision, Image and Deep Learning* (pp. 363–369). IEEE. https://doi.org/10.1109/CVIDL58838.2023.10167099

Geng, X., Su, Y., Cao, X., Li, H., & Liu, L. (2024). YOLOFM: An improved fire and smoke object detection algorithm based on YOLOv5n. *Scientific Reports*, *14*, 4543. https://doi.org/10.1038/s41598-024-55232-0

Giwa, O., & Benkrid, A. (2024). A new flame-based colour space for efficient fire detection. *IET Image Processing*, *18*(5), 1229–1244. https://doi.org/10.1049/ipr2.13022

Graham, T. L., & Roberts, D. J. (2000). Qualitative overview of some important factors affecting the egress of people in hotel fires. *International Journal of Hospitality Management*, *19*(1), 79–87. https://doi.org/10.1016/S0278-4319(99)00049-3

Jocher, G. (n.d.a). *Ultralytics/Ultralytics*. GitHub. https://github.com/ultralytics/ultralytics

Jocher, G. (n.d.b). *Ultralytics/Yolov5*. GitHub. https://github.com/ultralytics/yolov5

Johnston, J., Zeng, K., & Wu, N. (2022). An evaluation and embedded hardware implementation of YOLO for real-time wildfire detection. In *IEEE World AI IoT Congress* (pp. 138–144). IEEE. https://doi.org/10.1109/AIIoT54504.2022.9817206

Kim, J.-H., Kim, N., & Won, C. S. (2023). High-speed drone detection based on Yolo-V8. In *IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 1–2). IEEE. https://doi.org/10.1109/ICASSP49357.2023.10095516

Kobeissi, A., & Boulos, M. (2023). Innovative single-camera depth mapping for enhanced indoor fire detection and localization. In *4th International Multidisciplinary Conference on Engineering Technology* (pp. 278–282). IEEE. https://doi.org/10.1109/IMCET59736.2023.10368270

Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, Z., Li, Q., Cheng, M., Nie, W., Li, Y., Zhang, B., Liang, Y., Zhou, L., Xu, X., Chu, X., Wei, X., & Wei, X. (2022). *YOLOv6: A single-stage object detection framework for industrial applications*. arXiv. https://doi.org/10.48550/arXiv.2209.02976

Liang, D., Bui, T., & Wang, G. (2023). Fire and smoke detection method based on improved YOLOv5s. In *8th International Conference on Communication, Image and Signal Processing* (pp. 293–300). IEEE. https://doi.org/10.1109/CCISP59915.2023.10355820

Liu, J., Wang, P., & Song, G. (2019). Numerical study of kitchen fires in high-rise residential buildings. In *9th International Conference on Fire Science and Fire Protection Engineering* (pp. 1–4). IEEE. https://doi.org/10.1109/ICFSFPE48751.2019.9055772

Liu, S., Feng, J., Zhang, Q., & Peng, B. (2022). A real-time smoke and fire warning detection method based on an improved YOLOv5 model. In *5th International Conference on Pattern Recognition and Artificial Intelligence* (pp. 728–734). IEEE. https://doi.org/10.1109/PRAI55851.2022.9904105

Liu, W., Quijano, K., & Crawford, M. M. (2022). YOLOv5-tassel: Detecting tassels in RGB UAV imagery with improved YOLOv5 based on transfer learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 8085–8094. https://doi.org/10.1109/JSTARS.2022.3206399

Mahmoud, O., Saad, A., & Nazih, N. (2024). Deep learning-powered vision system for fire detection and localization in harsh environments. In *International Telecommunications Conference* (pp. 300–305). IEEE. https://doi.org/10.1109/ITC-Egypt61547.2024.10620511

Milke, J., & Zevotek, R. (2016). Analysis of the response of smoke detectors to smoldering fires and nuisance sources. *Fire Technology*, *52*, 1235–1253. https://doi.org/10.1007/s10694-015-0465-2

Mutar, A. F. (2019). Study fire detection based on color spaces. *Al-Mustansiriyah Journal of Science*, *29*(4), 93–99. https://doi.org/10.23851/mjs.v29i4.414

Ngai, M., Fu, E., Tam, A., Yang, A., & Ngai, G. (2022). Finding the signal from the smoke: A real-time, unattended fire prevention system using 3D CNNs. *Journal of Student Research*, *11*(3). https://doi.org/10.47611/jsrhs.v11i3.2981

Padilla, R., Netto, S. L., & da Silva, E. A. B. (2020). A survey on performance metrics for object-detection algorithms. In *International Conference on Systems, Signals and Image Processing* (pp. 237–242). IEEE. https://doi.org/10.1109/IWSSIP48289.2020.9145130

Rasbash, D. J. (1991). Major fire disasters involving flashover. *Fire Safety Journal*, *17*(2), 85–93. https://doi.org/10.1016/0379-7112(91)90030-3

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). *You Only Look Once: Unified, real-time object detection*. arXiv. https://doi.org/10.48550/arXiv.1506.02640

Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 6517–6525). IEEE. https://doi.org/10.1109/CVPR.2017.690

Redmon, J., & Farhadi, A. (2018). *YOLOv3: An incremental improvement*. arXiv. https://doi.org/10.48550/arXiv.1804.02767

Spearpoint, M., & Hopkin, C. (2020). A study of the time of day and room of fire origin for dwelling fires. *Fire Technology*, *56*, 1465–1485. https://doi.org/10.1007/s10694-019-00934-5

Vats, A., & Anastasiu, D. C. (2023). Enhancing retail checkout through video inpainting, YOLOv8 detection, and DeepSort tracking. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 5530–5537). IEEE. https://doi.org/10.1109/CVPRW59228.2023.00585

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). *YOLOv10: Real-time end-to-end object detection*. arXiv. https://doi.org/10.48550/arXiv.2405.14458

Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2022). *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*. arXiv. https://doi.org/10.48550/arXiv.2207.02696

Wang, C.-Y., Yeh, I.-H., & Liao, H.-Y. M. (2024). *YOLOv9: Learning what you want to learn using programmable gradient information*. arXiv. https://doi.org/10.48550/arXiv.2402.13616

Wang, S., & Wang, X. (2023). ES-YOLO: A new lightweight fire detection model. In T. Zhang & T. Yang (Eds.), *Third International Conference on Computer Vision and Data Mining* (Vol. 12511, pp. 347–355). SPIE. https://doi.org/10.1117/12.2660040

Wu, S., Li, Z., Li, S., Liu, Q., & Wu, W. (2023). Static gesture recognition algorithm based on improved YOLOv5s. *Electronics*, *12*(3), 596. https://doi.org/10.3390/electronics12030596

Zhang, D., & Chen, Y. (2024). Lightweight fire detection algorithm based on improved YOLOv5. *International Journal of Advanced Computer Science and Applications*, *15*(6), 809–816. https://doi.org/10.14569/IJACSA.2024.0150681

Zhao, Y., Wu, S., Wang, Y., Chen, H., Zhang, X., & Zhao, H. (2023). Fire detection algorithm based on an improved strategy of YOLOv5 and flame threshold segmentation. *Computers, Materials and Continua*, *75*(3), 5639–5657. https://doi.org/10.32604/cmc.2023.037829

Zulkiffli, Z. (2018, June 4). Lupa tutup api dapur faktor kebakaran meningkat [Forgetting to cover the kitchen fire increases the risk of fire]. *Berita Harian*. https://www.bharian.com.my/berita/kes/2018/06/433790/lupa-tutup-api-dapur-faktor-kebakaran-meningkat

Pertanika J. Sci. & Technol. 33 (6): 2833 - 2854 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Machine Learning Enabled IoT System for Agricultural Land Recommendation

Bayu Rima Aditya[1]*, Anindia Agusta Ken Nadila[1], Muhammad Qanit Al-Hijran[1], Muhammad Bintang Ramadhan[1] and Yudha Ginanjar[2]

[1]School of Applied Science, Telkom University, Bandung 40257, Indonesia
[2]Rastek Inovasi Digital, Bandung 40257, Indonesia

ABSTRACT

This study aims to develop an agricultural land recommendation system by integrating the Internet of Things (IoT) and machine learning (ML). IoT devices, including the JXBS-3001 soil sensor and Raspberry Pi Pico RP2040, collect real-time soil data, which is analyzed using the decision tree (DT) algorithm. The DT algorithm is chosen for its simplicity, efficiency, and interpretability over random forest (RF) and *k*-nearest neighbors (*k*-NN). It provides structured decision-making, faster training, and better handling of numerical data for parameters such as soil pH, nutrient content (NPK), moisture levels, and temperature. The findings show that the system provides accurate crop recommendations, helping farmers make informed decisions. The integration of IoT and ML enhances land assessment and optimizes agricultural productivity. Future improvements could include weather analysis and plant disease detection to further support decision-making.

Keywords: Agricultural, decision tree, Internet of Things, machine learning, recommendation system

ARTICLE INFO

Article history:
Received: 26 December 2024
Accepted: 06 June 2025
Published: 29 October 2025

DOI: https://doi.org/10.47836/pjst.33.6.18

E-mail addresses:
bayu@tass.telkomuniversity.ac.id (Bayu Rima Aditya)
anindiaagustaken@student.telkomuniversity.ac.id (Anindia Agusta Ken Nadila)
muhammadqanital@student.telkomuniversity.ac.id (Muhammad Qanit Al-Hijran)
muhammadbr@student.telkomuniversity.ac.id (Muhammad Bintang Ramadhan)
yginanjar@gmail.com (Yudha Ginanjar)
* Corresponding author

INTRODUCTION

Agricultural management refers to the oversight of farming practices, which play a vital role in the agricultural industry by directly influencing the primary goal of agriculture: ensuring a sustainable food supply. According to the Food and Agriculture Organization of the United Nations (FAO) (2021), recent years have seen increasing demands for agricultural practices to deliver at least 70% of global food production annually. This requirement

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

is driven by the rapid growth of the global population, which is projected to reach approximately 9.6 billion by 2050 (FAO, 2021). Adding to the challenge, the availability of arable land continues to decline each year, further intensifying the need for effective agricultural management (Wang, 2022).

Indonesia is one of the countries with immense agricultural potential, supported by abundant natural resources and a favorable climate. However, the implementation of agricultural recommendation systems in farming activities remains underdeveloped (Iskandar, 2021). This is concerning, as limited knowledge among farmers about critical factors in agricultural decision-making, such as identifying the optimal planting season, evaluating land performance, or determining the most suitable crops for specific periods, continues to hinder productivity (Iskandar, 2021; Yundari, 2017). If left unaddressed, these challenges could lead to suboptimal agricultural practices, ultimately failing to meet Indonesia's growing demand for agricultural output. This condition indicates the importance of data-based decision-making to support farmers in choosing the most suitable crops (Priyadharshini et al., 2021; Ridoy et al., 2024).

To promote sustainability in Indonesia’s agricultural industry, innovation is essential across various aspects of farming practices (Rajeswari et al., 2018). One key area for innovation is decision-making, which can be significantly enhanced by developing agricultural recommendation systems powered by ML (Singh et al., 2024). ML, a branch of artificial intelligence, focuses on developing methods that can learn from past data (Lestari et al., 2023). ML has emerged as a transformative tool across many industries, including agriculture, by providing advanced, data-driven solutions to boost efficiency and productivity (Musanase et al., 2023; Ridoy et al., 2024). In agriculture, ML leverages data from sensors and other sources to analyze land conditions, predict yields, and optimize resource use (Akintuyi, 2024). This technology enables farmers to make more accurate, data-informed decisions while mitigating risks associated with weather fluctuations and soil variability. By integrating ML, agriculture can evolve toward smarter, more sustainable methods, paving the way for a revolutionary shift in farming practices.

The main objective of this study is to develop a technology-driven agricultural land recommendation system. This system integrates IoT and ML to provide farmers with precise recommendations for selecting agricultural land suitable for their desired crops. IoT is a concept of integration that enables any object to send data through network modernization (Kiruthika & Karthika, 2023). In this study, the system acts as an intermediary between farmers and the system, enabling farmers to input desired parameters directly through the interface. These parameters are then transmitted to the ML system for further processing. Various ML techniques can be applied for recommendations, such as DT, Naive Bayes (NB), RF, rotation forest, *k*-NN, artificial neural networks (ANN), and support vector machine (SVM), among others. Each method has its unique strengths and limitations. DT,

for instance, is chosen for its high accuracy and ability to trace each decision step taken by the algorithm. Additionally, DT can seamlessly integrate with database systems, offer strong accuracy, and identify unexpected data combinations (Sihombing & Arsani, 2021). It also effectively represents patterns, knowledge, or insights in a decision-tree format (Khoeri & Mulyana, 2021).

By integrating IoT and ML to facilitate farmers with the system and employing ML algorithms, specifically the DT method, the proposed system aims to conduct in-depth analyses of collected data. This process ultimately delivers tailored recommendations for the most suitable crop commodities for specific agricultural lands. Therefore, this study aims to design and develop an agricultural land recommendation system based on IoT and ML technologies.

RELATED WORKS

IoT-based land-monitoring systems have attracted significant attention in recent years, enabling data-driven decision-making in agriculture. IoT enables real-time data acquisition from diverse sensors, including moisture, pH, and nutrient sensors, to assess soil conditions (Aditya et al., 2024; Gaikwad et al., 2021; García et al., 2020; Ramson et al., 2021). However, existing IoT-based systems often lack intelligent decision-making capabilities, necessitating the integration of ML models to enhance agricultural recommendations.

Various ML techniques have been explored in crop recommendation, with studies utilizing RF, SVM, and ANN to improve predictive accuracy (Lokhande et al., 2022; Qiu et al., 2021; Vemulapalli et al., 2024). The study by Modi et al. (2021) highlighted the risks associated with improper crop selection, including reduced yields and severe issues, such as rising suicide rates among farmers. To mitigate these challenges, they proposed a crop recommendation system based on the SVM algorithm. This system analyzed soil parameters, including nitrogen (N), phosphorus (P), potassium (K), pH, moisture levels, rainfall, and temperature. Achieving an accuracy rate of 97%, the system provided farmers with precise recommendations for selecting the most suitable and productive crops, thereby reducing errors in crop selection and increasing agricultural output. The suggested approach does not focus on real-time processing.

Several studies have applied DT algorithms for agricultural recommendations due to their efficiency and interpretability (Khoeri & Mulyana, 2021; Sihombing & Arsani, 2021). While DT offers advantages in terms of fast training and explainability, its performance in handling complex feature interactions remains a challenge.

In a separate study by Islam et al. (2023), the authors present an IoT-based solution that integrates ML techniques to optimize crop production through real-time soil monitoring. The system uses various sensors to measure soil nutrients, moisture, temperature, and humidity, and sends the data to a server for analysis. Using ML algorithms, the system

provides customized crop recommendations and fertilizer-use guidelines. However, this model cannot assess the percentage accuracy of its crop recommendations.

Another notable study by Rao et al. (2022) explored the impact of nutrient deficiencies and incorrect crop selection on agricultural outcomes. The research compared three ML algorithms — *k*-NN, DT, and RF classifier — using Gini and entropy criteria. The findings revealed that the RF classifier achieved the highest accuracy of 99.32%, while *k*-NN had the lowest accuracy of 97.04%. The DT algorithm performed moderately, with the Gini criterion yielding better results than entropy. However, this study still has limitations, particularly the lack of integration between crop prediction and IoT technology, which could enable real-time data collection to improve the accuracy and relevance of recommendations.

Recent advancements in agricultural technology, particularly the integration of IoT and ML, enable farmers to collect and analyze data in real time (Senapaty et al., 2023). In a separate study by Islam et al. (2023), the authors present an IoT-based solution that integrates ML techniques to optimize crop production through real-time soil monitoring. The system uses various sensors to measure soil nutrients, moisture, temperature, and humidity, and sends the data to a server for analysis. Using ML algorithms, the system provides customized crop recommendations and fertilizer-use guidelines. The suggested approach does not consider portability for farmers, as it is not designed in an integrated device that can be easily transported to various locations.

This study addresses the gap in the literature by integrating IoT and DT-based ML models to provide real-time, resource-efficient, and interpretable crop recommendations. Unlike previous works that rely on static datasets, this research incorporates real-time sensor data via IoT devices, enabling dynamic adjustments based on environmental conditions. Additionally, the proposed system optimizes DT hyperparameters to balance accuracy and computational efficiency, making it more suitable for deployment in low-power agricultural devices.

MATERIALS AND METHODS

IoT Devices and Sensors

This study uses three sensor types: a 3.5-inch thin-film transistor liquid-crystal display (TFT LCD) module (China), a Raspberry Pi Pico RP2040 (Raspberry Pi Foundation, United Kingdom), and a JXBS-3001 soil sensor (Weihai Jingxun Changtong Electronic Technology Co., Ltd., China). The circuit diagrams in Figures 1 and 2 present the configuration and specifications of the proposed IoT system components.

a. 3.5-inch TFT LCD

   TFT LCD is a type of display screen that uses thin-film transistors to control each pixel individually, resulting in sharper and more responsive images. In this system, a TFT

LCD displays the results of soil sensor data analysis. This component is a 3.5-inch (77.98 mm × 43.94 mm) touch display with a resolution of 320 × 240 pixels, offering sharp, detailed graphics for visual interfaces. It is equipped with a resistive touchscreen, allowing users to interact directly with the device. The display uses an ILI9341-16 controller that supports a 16-bit data interface. It is powered by the Universal TFT Library from Rinky-Dink Electronics, which simplifies programming across a variety of Arduino platforms. The display is compatible with the Arduino Mega and Arduino Due microcontrollers, both of which offer sufficient memory and I/O pins to handle the high-resolution data from the display. However, it does not support Arduino Uno with an 8-bit serial interface due to data transmission limitations.

b. Raspberry Pi Pico RP2040

The Raspberry Pi Pico RP2040 module serves as the primary microcontroller, facilitating communication between sensors and other system components. This printed circuit board (PCB) supports the SmartMatrix Arduino Library, ensuring optimal compatibility for a variety of visual applications. The Raspberry Pi Pico RP2040 delivers strong data processing capabilities, bolstered by Wireless Fidelity (Wi-Fi) connectivity. This enables the system not only to receive input from sensors but also to process the data using ML algorithms and transmit or receive information wirelessly. With good energy efficiency, this module also supports SmartMatrix HUB75 connectivity, specifically designed for red, green, and blue (RGB) matrix light-emitting diodes (LEDs). The pinouts available on this module enable high-precision, high-speed LED control, ensuring effective synchronization of visual elements on LED displays.

c. JXBS-3001 Soil Sensor

The JXBS-3001 soil pH sensor is an important component in the soil monitoring system, designed to integrate seamlessly with the Raspberry Pi Pico RP2040 via RS-485 communication. Powered by a 12-24V DC supply and utilizing the Modbus protocol, this sensor is capable of measuring soil parameters, including NPK, total dissolved solids (TDS), moisture, temperature, and pH, with a precision of ±0.3. The sensor's real-time data is transmitted to the Raspberry Pi Pico RP2040 for processing and displayed via an LED interface or a TFT screen. The integration of the JXBS-3001 soil sensor with the Raspberry Pi Pico RP2040 enables the system to provide optimal recommendations on crop land suitability. Supported by the microcontroller's rapid processing capabilities and wireless connectivity, the system ensures efficiency, stability, and reliability in field applications.

d. Direct current to direct current (DC-to-DC) step-up converter

The DC-to-DC step-up converter, also known as a boost converter, is a device that increases the input voltage to a higher output level. It is particularly useful when a

device requires a higher voltage than the available power source, such as boosting low voltage from a battery to operate specific electronic components. The boost converter uses inductive charging to efficiently raise the voltage, making it an ideal choice for portable or renewable energy systems. In microcontroller-based sensor and control applications, the step-up converter plays a critical role in maintaining optimal performance by ensuring adequate voltage levels for various system components, even when operating on limited power sources.

e. 3.7 V battery

The battery serves as the primary or backup power source in electronic systems, providing stable and continuous energy to ensure smooth operations. In applications involving the Raspberry Pi Pico RP2040 or similar microcontrollers, the battery enables portable, independent operation, eliminating the need for a constant connection to an external power source. The battery type and capacity are tailored to meet the device's power requirements and the desired operational duration.

f. RS-485

RS-485 is a serial communication protocol widely used to connect devices that require long-distance communication in industrial networks or control systems. Its advantages include its ability to transmit data over long cable lengths with minimal interference and its support for multipoint communication. In systems with sensors and microcontrollers, RS-485 is often used to ensure stable, efficient device-to-device communication, even in challenging environments.

The architectural diagram illustrates a structured system integrating the JXBS-3001 soil sensor with the Raspberry Pi Pico RP2040, employing Wi-Fi communication to send data to an external application programming interface (API). The JXBS-3001 soil sensor measures soil parameters such as moisture, pH, and nutrients, transmitting data through the RS-485 protocol to the Raspberry Pi Pico RP2040, which serves as the system's data processing hub. RS-485 ensures reliable communication over long distances with minimal interference. The Raspberry Pi Pico RP2040 processes sensor data and directs it to two main outputs: first is real-time visualization on a 3.5-inch TFT LCD for immediate user feedback, and second is transmission via Wi-Fi using the Transmission Control Protocol/ Internet Protocol (TCP/IP) to an external API for remote monitoring and integration with other applications. TCP/IP is a network protocol that standardizes communication on the internet and local networks. The system is powered by a 3.7 V battery connected to a DC-to-DC step-up converter to ensure a stable and sufficient voltage supply. For charging, there is a charger controller module that is connected to the recreational vehicle (RV) DC charger, managing the battery recharge to keep it charged and allowing the system to operate

for extended periods without interruption. Overall, this integrated soil monitoring system provides real-time data visualization, remote data transmission, and optimized power efficiency with a rechargeable battery, making it ideal for field applications independent of constant external power sources.

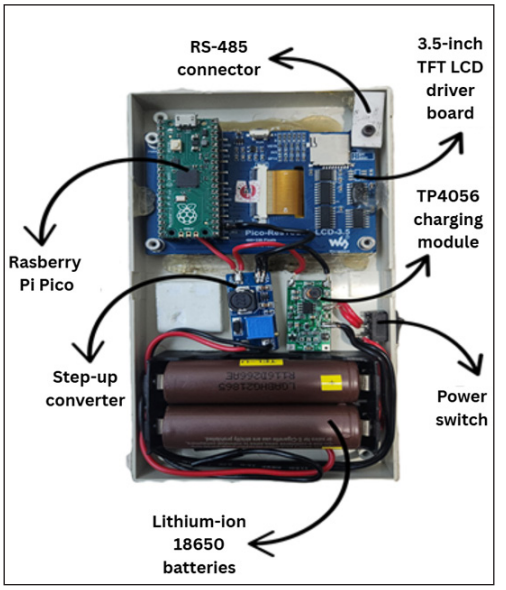


Figure 1. Block diagram of a portable device
Note. TFT LCD = Thin-film-transistor liquid-crystal display; RS-485 = Recommended standard 485

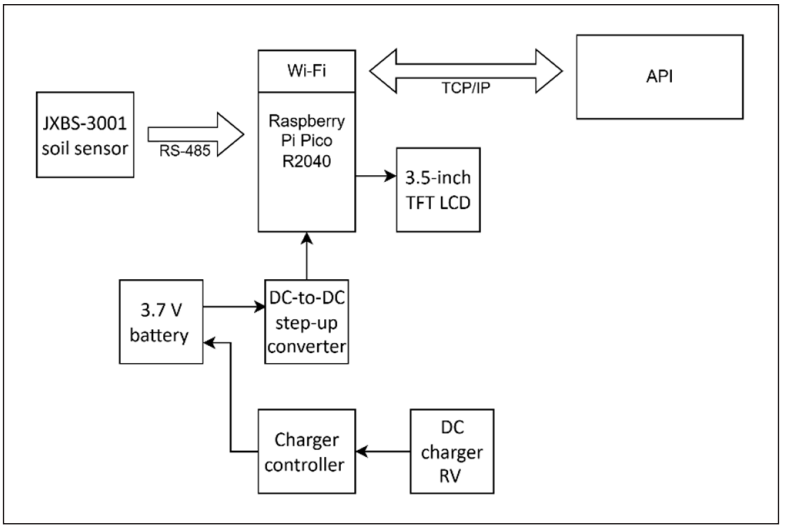


Figure 2. Schematic of an embedded system
Note. TFT LCD = Thin-film-transistor liquid-crystal display; TCP = Transmission Control Protocol; IP = Internet Protocol; API = Application programming interface; DC = Direct current

Data Transmission Process

In the described communication system, TCP/IP serves as the foundational network protocol, enabling data exchange between devices and servers via Hypertext Transfer Protocol (HTTP), a protocol used for data exchange over the internet. The process begins with the device sending a request to the server using the POST method. Data is structured in JavaScript Object Notation (JSON) format and included in the HTTP body. TCP ensures reliable delivery and correct sequencing of data, while IP handles addressing and routing across the network.

Once the server receives the data, it processes it based on the predefined application logic. The server then responds to the device in JSON format via HTTP. TCP connection ensures that the response data maintains integrity and proper sequencing, while IP manages addressing and routing to the correct destination on the network. The HTTP protocol governs the formatting and structuring of data communication, ensuring it can be understood and correctly processed.

The device then receives the server's response and displays the information to the user. This entire communication process, from sending a request to receiving a response, is supported by TCP/IP, which provides the basic mechanism for data transmission. Meanwhile, HTTP offers an application-level protocol to structure and format the transmitted data efficiently. This communication cycle highlights the integral implementation of TCP/IP in supporting efficient and structured HTTP-based communications.

ML Methods

a. DT algorithm classification report

The *croprekomendasi.csv* dataset has been analyzed and cleaned to improve data quality for crop recommendations based on environmental factors. The initial analysis showed no missing values, but several features, such as illumination (illum), rainfall (rain), and wind speed (wind), contained extreme values or uneven distributions. For instance, the illumination value reached 216, while 75% of the data had values below 9.86. Similarly, rainfall and wind speed had many zero values, as well as some extremely high values. To address this issue, the interquartile range (IQR) method was applied to detect and remove outliers. As a result, 998 data points (32.6%) were removed from the total 3,063 records, leaving 2,065 cleaned data points.

The DT is chosen over RF and *k*-NN due to its simplicity, efficiency, and ease of interpretation. Its clear tree structure allows for more transparent analysis than RF, which consists of multiple trees and is difficult to trace, and *k*-NN, which does not build an explicit model. In terms of efficiency, the DT trains are faster than RF, which requires multiple iterations to construct various DTs, and it also predicts faster than

k-NN, which needs to compare each new data point with the entire training dataset. Additionally, the DT does not rely on distance metrics like *k*-NN, making it more suitable for data with numerical and categorical attributes that have different scales. This algorithm is also easier to optimize as it does not require extensive parameter tuning, such as determining the number of trees in RF or selecting the optimal *k* value in *k*-NN. Another advantage is its ability to handle hierarchical data, such as classifying crops based on environmental factors, which is difficult for k-NN, which relies solely on data proximity without understanding rule-based hierarchies. With these advantages, the DT becomes a more suitable choice for classifying sensor data and processed data.

The backend service will store the accumulated data in the database. The stored data includes both raw and processed sensor data. A popular machine-learning technique for classification tasks is the DT algorithm, which is based on predictive modeling. In this context, the algorithm partitions a dataset into groups based on specific characteristics relevant to the target categories. The DT starts with a root node and branches into smaller nodes, where each branch represents a decision or classification based on attributes. To produce leaf nodes that reflect the most appropriate classification choices, the primary objective is to optimize homogeneity within each data subset at each branch.

Choosing the most informative attribute to divide the data into smaller subsets is the first step in the classification process. This attribute is selected based on criteria such as information gain (IG) or the Gini index, which measure how well it can classify the data. For instance, soil pH or moisture might be critical attributes for predicting suitable crops, as they have significant discriminative strength for separating plant types based on their needs.

The DT divides the data iteratively during the learning process, with each new node representing a classification decision based on the tested attribute's value. Once the tree is fully developed, new data can be classified by following a series of binary decisions that guide the data from the root node to the leaf nodes, where a final classification or recommendation is provided.

b. DT ML algorithm for recommendations

To make crop recommendations, the DT algorithm divides the dataset into smaller subsets based on relevant characteristics, such as soil pH, nutrient content (NPK), moisture levels, and temperature. The algorithm maps the relationships between these parameters and the crops best suited to the conditions, using a DT structure derived from a series of if-then-else conditions. Each node in the tree represents a test on a specific feature (e.g., soil pH), the branches show the test results, and the final leaves offer the crop recommendation.

The tree-building process involves techniques that either maximize IG or minimize the Gini index (GI). At each step, the algorithm selects the most significant attribute to split the data and create a new node. The IG is calculated using an entropy formula:

$$Entrophy\ (S) = -\sum_{i=1}^{n} P_i\ P_i$$

In this context, P_i represents the proportion of the class i within subset S. The IG is then calculated as follows:

$$IG(T, A) = Entropy\ (T) - \sum_{v \in Values\ (A)} \frac{|T_v|}{|T|} Entrophy\ (T_v)$$

where T is the entire dataset, A is the attribute being tested, and T_v is the subset of data where attribute A has value v. A higher IG indicates that the attribute is more effective in dividing the data, making it a higher-priority candidate for a DT node.

Once the DT is constructed, new data (e.g., soil sensor data) is analyzed by starting at the root node and traversing branches based on the conditions the data meet until a leaf node is reached that provides a crop recommendation. This algorithm effectively recommends the best crop varieties based on agricultural conditions, including soil pH, moisture levels, and temperature. Its strength lies in its ability to handle multidimensional data and produce decisions that farmers can understand and use with ease.

Proposed Framework

The proposed framework integrates IoT sensors, data transmission, and ML to enhance decision-making processes in agriculture. The system employs various soil sensors (pH, NPK, TDS, moisture, and temperature) installed in agricultural fields to measure soil conditions in real-time. The proposed framework is illustrated in Figure 3. These sensors are connected to portable devices used by farmers, which serve as data-collection hubs. The portable devices are equipped with Raspberry Pi Pico RP2040, responsible for gathering data from sensors and transmitting it to remote servers via the Internet using TCP/IP and HTTP protocols.

ML algorithms are used to process the data when it arrives at the server. These algorithms analyze soil parameters and generate personalized crop recommendations tailored to the current soil conditions. Appropriate crop varieties, ideal planting times, and accurate fertilizing techniques are among the suggestions. After that, the analysis

results are then sent back to the farmer's portable device, where they may see the data and suggestions right on the screen.

Additionally, the system provides continuous soil condition monitoring, enabling farmers to make data-driven decisions that enhance productivity and efficiency. To accommodate an increasing number of users and agricultural lands, the framework is designed with efficient server management, including optimized database handling, caching mechanisms, and load balancing strategies to maintain performance as demand grows. By combining IoT and ML technologies with a well-structured server-based infrastructure, the proposed method aims to optimize land use and maximize agricultural yields with precision and reliability.

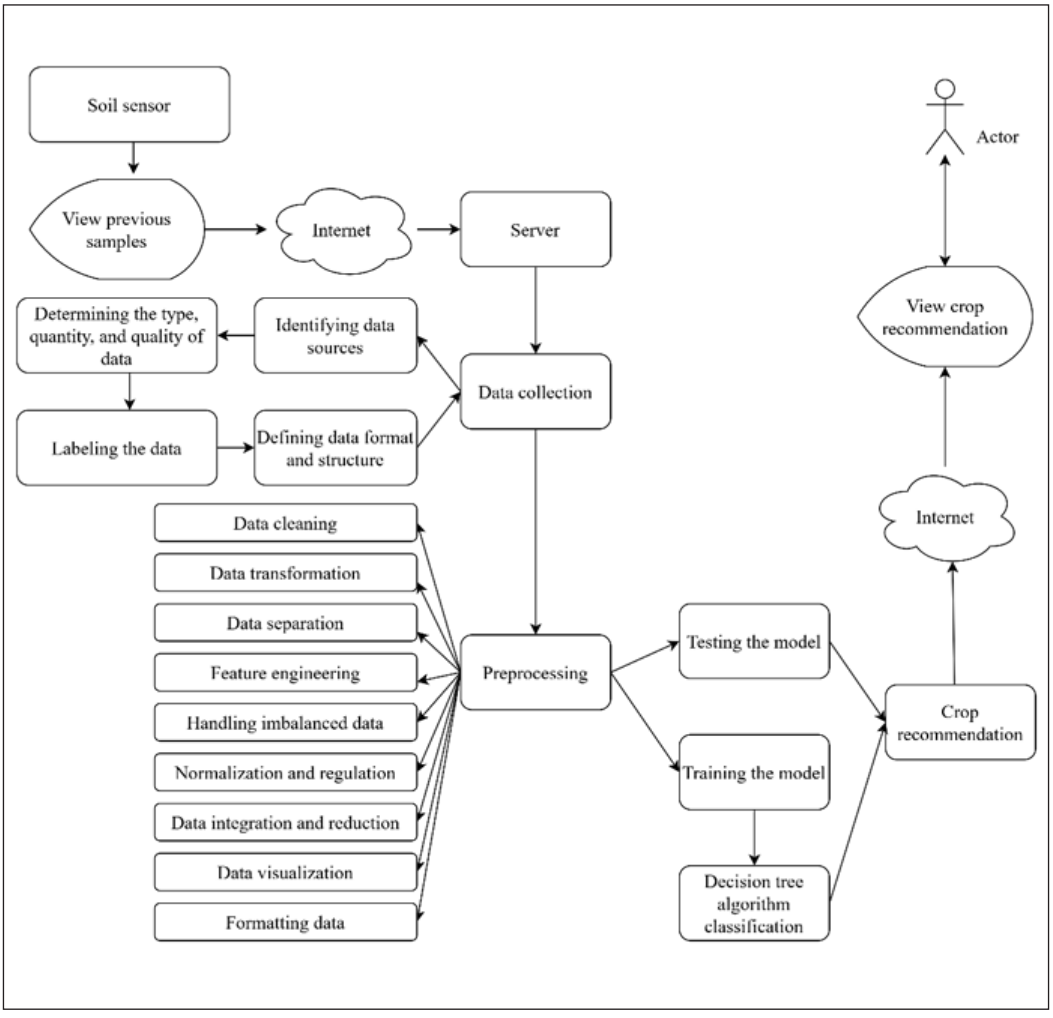


Figure 3. Framework for intelligent crop recommendation

RESULTS

Experimental Setup

The system comprises a single sensor that measures critical soil parameters, such as pH, moisture, and nutrient levels. This sensor connects to a microcontroller unit (MCU) via serial input/output (I/O), which manages data acquisition from the sensor. The MCU serves as the primary processing hub, handling the collection, processing, and transmission of sensor data to other system components. Figure 4 illustrates the system block diagram of the proposed system.

A display module is incorporated to provide real-time visualization of the sensor data, enabling users to monitor soil conditions instantly. Additionally, a Wi-Fi module facilitates wireless data transmission, linking the device to a server or cloud system via TCP/IP and HTTP protocols. The server can further analyze the transmitted data using ML algorithms to generate recommendations for crops best suited to the current soil conditions.

The system also includes a storage component to store data locally for historical records or in-depth analysis. A power supply ensures all components remain operational, supporting uninterrupted functionality. This experimental setup is designed to continuously collect soil data, enabling ML-driven analyses that deliver actionable crop recommendations that can be directly accessed via the display.

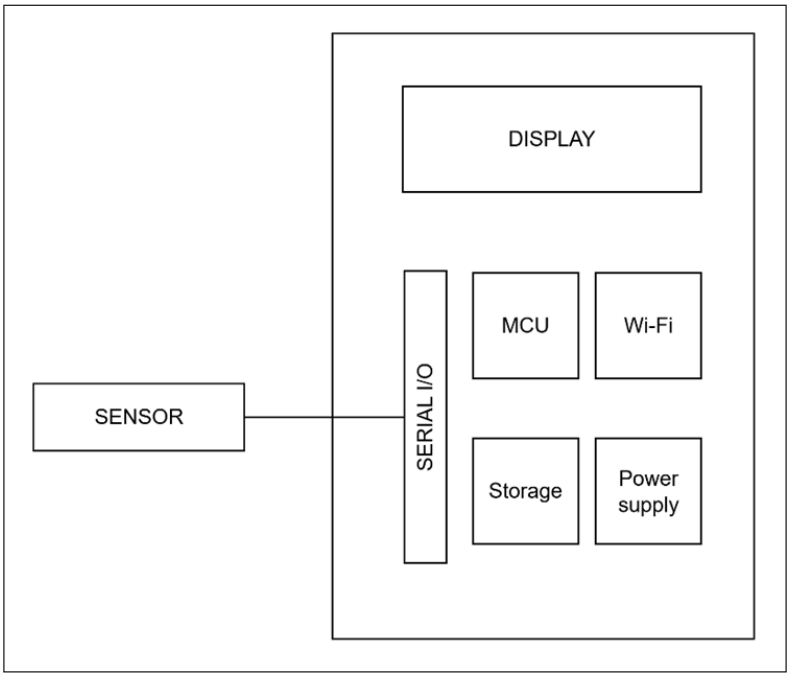


Figure 4. System block diagram
Note. SERIAL I/O = Serial input/output; MCU = Microcontroller unit; Wi-Fi = Wireless Fidelity

Data Collection and Preprocessing

The collected data originates from real-time environmental and soil monitoring, serving as input for training, validating, and testing the ML model. High-quality, relevant data is crucial to developing an accurate and effective model. The data collection process follows these steps:

1. Identifying data sources
2. Determining the type, quantity, and quality of data
3. Labeling the data
4. Defining data format and structure

Data preprocessing is an essential step in evaluating the model's performance and ensuring its reliability. Therefore, data cleaning, data transformation, data separation, feature engineering, handling imbalanced data, normalization, and regularization, data integration and reduction, data visualization, and organizing data into a format suitable for training and testing the model are carried out.

Summary Statistics of a Feature in ML

Tables 1 and 2 present a statistical summary of a specific feature within the ML dataset. They include the minimum, maximum, and average values of all sensor parameters observed for the crops under study, specifically corn and shallots.

Table 1
Statistical summary of corn

Parameter	Min	Mean	Max
N (mg/kg)	0.00	6.32	17.00
P (mg/kg)	29.40	59.96	85.64
K (mg/kg)	23.00	52.38	78.18
Soil temperature (°C)	9.84	26.11	38.63
Humidity (%)	46.62	86.21	99.90
pH	22.80	55.13	69.50
Conductivity (µS/cm)	97.20	213.9	267.0
Soil humidity (%)	12.88	25.44	46.17
Temperature (°C)	18.30	26.09	35.60
TDS (ppm)	48.40	106.70	133.18
Salinity (ppt)	53.20	117.20	146.30

Note. Min = Minimum; Max = Maximum; N = Nitrogen; P = Phosphorus; K = Potassium; pH = Potential of hydrogen; TDS = Total dissolved solids

Table 2
Statistical summary of shallots

Parameter	Min	Mean	Max
N (mg/kg)	0.00	1.53	69.20
P (mg/kg)	5.03	32.73	206.70
K (mg/kg)	2.70	24.90	200.10
Soil temperature (°C)	16.81	24.84	35.70
Humidity (%)	42.51	80.01	99.90
pH	29.83	48.65	90.00
Conductivity (µS/cm)	38.06	157.53	518.11
Soil humidity (%)	0.58	9.68	33.43
Temperature (°C)	18.30	26.09	35.60
TDS (ppm)	19.22	78.53	258.80
Salinity (ppt)	21.27	86.17	284.50

Note. Min = Minimum; Max = Maximum; N = Nitrogen; P = Phosphorus; K = Potassium; pH = Potential of hydrogen; TDS = Total dissolved solids

DT Algorithm Classification Report

Figure 5 illustrates the classification report generated by the DT algorithm. The table evaluates each crop, providing metrics such as precision, recall, F1 score, and support. Precision represents the proportion of correctly identified positive cases, while recall measures the proportion of actual positive instances that were accurately classified. The F1 score is a harmonic mean of precision and recall, offering a balanced metric for performance evaluation. Support indicates the number of samples used to calculate each metric. Additionally, the report includes macro averages and weighted averages to reflect performance across multiple classes.

```
DecisionTrees's Accuracy is:  98.36779107725789
              precision    recall  f1-score   support

           0       0.99      0.98      0.99       540
           1       0.97      0.99      0.98       379

    accuracy                           0.98       919
   macro avg       0.98      0.98      0.98       919
weighted avg       0.98      0.98      0.98       919
```

Figure 5. Classification report for a decision tree model
Note. avg = Average

Accuracy of Crop Recommendation Models

Table 3
Accuracy of crop recommendation

No.	Model	Accuracy
1	Random forest	0.99
2	SVM	0.90
3	Naïve Bayes	0.91
4	Decision tree	0.98
5	XGBoost	0.99

Note. SVM = Support vector machine

The XGBoost, known for its ensemble approach to combining DT models, and the RF, which similarly aggregates multiple DTs, generally outperform standalone DT algorithms. These ensemble methods leverage the collective insights of multiple trees, enhancing their overall predictive accuracy. In contrast, individual DTs may underutilize some features during training. The significance of each attribute in these models accounts for all features except those excluded during training, which can limit the performance of a single DT.

The high accuracy achieved by the crop recommendation model reflects the rigorous optimization techniques applied throughout its development. Careful feature selection ensured that only the most relevant soil and environmental parameters were included, with redundant or less impactful features removed based on correlation analysis. This step allowed the model to focus on the most meaningful data, improving its predictive performance. Furthermore, extensive hyperparameter tuning using techniques such as grid search and Bayesian optimization was conducted to improve the model's learning.

In addition, the ensemble nature of XGBoost and RF enabled the model to capture complex relationships in the data, reducing bias and variance and improving generalization. The dataset itself was carefully curated, cleaned, and preprocessed to ensure high data quality, minimizing the risk of noise and inconsistencies. To further validate the model's robustness, cross-validation techniques were employed, ensuring that the accuracy remained consistent across different subsets of data.

By integrating these advanced methodologies, the crop recommendation model effectively leverages the strengths of ensemble learning and rigorous data processing, leading to its impressive accuracy. Table 3 presents the detailed accuracy results of the model.

The DT algorithm was chosen as the main model in this system not only because of its high accuracy, but also because it has a low absolute error value. This low error is significant in crop recommendations, where a small error can impact the decision taken by the farmer, such as the selection of the wrong crop. In addition, DT offers a good level of interpretability, which allows the system to provide clear and easy-to-understand recommendations. The DT structure used in this model makes the results explainable in a more transparent way. Therefore, although ensemble methods may be considered in the future to improve the accuracy of the system, DT remains the top choice due to its balance between good accuracy and easier interpretability.

Dimensionality Reduction with Variable Feature Issue

In this context, XGBoost and DT models demonstrate minimal errors, largely due to their ability to handle non-linear relationships and effectively manage data complexity. XGBoost employs iterative boosting techniques to refine model performance, while DTs adaptively learn the structural intricacies of data. DTs are particularly effective in capturing non-linear relationships between features and target variables. By adaptively partitioning the feature space, the DT can modify the model for training data, resulting in accurate predictions. Table 4 presents the absolute error of the crop recommendation model.

Table 4
Absolute error of crop recommendation

No.	Model	Absolute error
1	Random forest	0.009793
2	SVM	0.097933
3	Naïve Bayes	0.091404
4	Decision tree	0.013058
5	XGBoost	0.008705

Note. SVM = Support vector machine

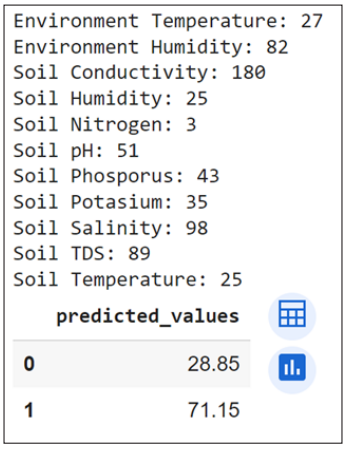


Figure 6. Soil sensor data and predicted values

Crop Recommendation Results

Figure 6 illustrates the crop recommendation results obtained after applying the DT ML algorithm. In the results, crops are represented with codes: code 0 indicates shallots, while code 1 represents corn. This classification identifies the most suitable crop for a given soil condition.

IoT-ML Enabled Agriculture Platform

This study introduces a crop recommendation model for corn and shallots, based on an ML architecture with various parameter configurations. The DT algorithm was employed to process and manage soil parameter data. The model's performance was evaluated using multiple metrics, including accuracy, loss, precision, recall, F1 score, and confusion matrix. Experimental results demonstrated that the DT algorithm provides highly accurate predictions for crop-to-land suitability. This is evidenced by the high accuracy achieved on test data (98%), supporting its potential for future applications. In summary, compared to conventional techniques, the ML-based crop and soil suitability recommendation model with the DT model produces better crop suitability prediction outcomes. A preview of the suggestion tool's user interface (UI) is displayed in Figures 7-10.



Figure 7. Set sampling parameter interface

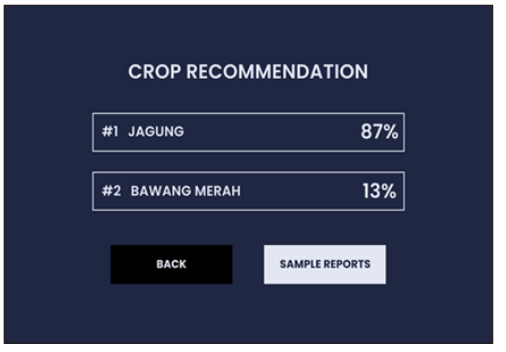


Figure 8. Crop recommendation interface



Figure 9. View sample interface



Figure 10. Proposed device

DISCUSSION

This study suggests a method based on ML and the IoT to collect data on soil parameters and suggest appropriate crops depending on the information gathered. The JXBS-3001 soil sensor, which measures soil characteristics like pH, moisture, and nutrient concentrations (NPK and TDS), is one of the main components. A Raspberry Pi Pico RP2040 serves as the data collection hub, connecting to the sensor via the RS-485 protocol. The TCP/IP and HTTP protocols are used to send sensor data to a server, enabling remote monitoring and real-time data processing. A portable device with a 3.5-inch TFT LCD screen provides farmers with analyzed soil data, facilitating data-driven decision-making in the field.

On the server, data from the sensors is processed using ML models, notably the DT algorithm, to recommend crops suited to the soil conditions. This algorithm functions as a classification tool, utilizing the measured soil parameters as inputs. It categorizes data into groups based on decision rules generated from the training dataset. The classification process helps identify the most suitable crop for specific soil conditions by analyzing soil pH, temperature, and moisture levels. Using IG and entropy, the algorithm selects the most effective attributes to split the data, resulting in accurate and reliable recommendations.

The primary advantage of this system lies in its use of the DT algorithm for crop recommendation. This algorithm enables the system to develop models that handle non-linear data, allowing precise predictions of the most suitable crops under complex soil conditions. Furthermore, the system's ability to process large volumes of data through a server integrated with ML models ensures efficiency in decision-making, a feat that traditional methods often struggle to achieve.

The integration of IoT and ML within this framework also highlights the benefits of automation and remote monitoring in agriculture. This system not only reduces farmers' workload but also enhances land management accuracy, particularly in selecting optimal crops. With real-time data collection, farmers can make informed decisions based on algorithm-driven analyses while continuously monitoring soil and crop conditions through a portable device equipped with a visual display.

Additionally, utilizing the server as a processing platform supports system scalability, allowing ML models to improve as more data is collected from sensors. This scalability opens the door to further development, such as integrating plant disease analysis modules or improving yield predictions. By employing this technology, the system can significantly increase agricultural productivity while promoting environmental sustainability by optimizing fertilization and irrigation using accurate data.

To ensure that the developed system can be used effectively by farmers, usability testing was conducted by considering three main aspects: efficiency, effectiveness, and system usability scale (SUS). The test results showed that the system achieved an efficiency and effectiveness level of 100%, indicating that users can complete tasks without obstacles and in an efficient time. The SUS score of 81 indicates a very positive user assessment, which classifies the system into the “excellent” category. The overall average of the three indicators reached 93%, confirming that the user interface design, designed with farmers' needs in mind, has resulted in an optimal usage experience in the agricultural context.

Finally, collaboration between researchers, agricultural practitioners, and technology developers can strengthen the implementation of these solutions in the field. A participatory approach to research and development will ensure that the resulting systems are relevant and adaptable to the real needs of farmers, which, in turn, will increase the adoption of technology within the agricultural sector. Through these steps, future research could make significant contributions to the development of more efficient and sustainable agricultural practices.

CONCLUSION

In this study, a combination of an IoT-based system and ML was successfully developed to collect soil parameter data and provide suitable crop recommendations. The use of the JXBS-3001 soil sensor and the Raspberry Pi Pico RP2040 device allowed for the accurate

collection of soil condition data, which was then processed using a DT algorithm to generate data-driven recommendations.

The primary strength of this system lies in its ability to manage complex data and provide smarter solutions for agricultural decision-making. Additionally, the integration of IoT and ML not only enhances land management efficiency but also enables better real-time monitoring of soil and crop conditions.

Through this approach, farmers can make more informed decisions regarding crop selection and farming practices, ultimately contributing to increased productivity and sustainability in the agricultural sector. This research also opens up opportunities for further development, including the integration of weather analysis modules and plant disease detection, which could further strengthen the system in addressing the challenges of modern agriculture. Overall, the results of this study highlight the great potential of revolutionizing agricultural practices with a more technology- and data-driven approach.

FUTURE RESEARCH

In future work, scalability testing of the system is needed to evaluate its performance in the face of increasing data loads and higher user demands. These tests include the system's ability to handle large data volumes, the reliability of data transmission on networks with limited bandwidth, and the server's capacity to handle multiple simultaneous requests. All these tests aim to ensure the system can function efficiently and reliably on a larger scale and under varied field conditions.

Overall, the proposed system demonstrates significant potential in revolutionizing modern agriculture by introducing a smarter and more efficient approach. The integration of IoT with ML algorithms offers a comprehensive solution that enhances crop productivity, optimizes resource utilization, and reduces uncertainty in land management. Future work could include further development to support weather-based farming, plant growth pattern recognition, and more advanced pest detection systems.

This study outlines several future research directions that could expand the effectiveness of the proposed system. First, the development of more advanced ML models, such as ensemble algorithms or deep learning, could be applied to improve the accuracy of crop recommendation predictions. This approach would allow the system to capture more complex patterns in the data collected from sensors, while also enhancing the model's resilience to data variability.

Second, there might be further advantages to incorporating weather monitoring technology into the system. Farmers may gain a more comprehensive understanding of the environmental factors that could impact crop development by utilizing weather prediction modules. This would enable timelier, data-driven decision-making regarding planting, fertilization, and irrigation.

Furthermore, research could focus on developing early detection systems for plant diseases. By utilizing visual data or imagery collected from drones or cameras, pattern recognition algorithms could be employed to detect early signs of plant diseases, allowing for prompt and efficient intervention. This would improve crop resilience and reduce yield losses due to pest and disease outbreaks.

Additionally, more studies should examine environmental sustainability in relation to better farming methods. Research on the application of technology that facilitates precision farming methods could be combined to optimize land management while reducing water and fertilizer consumption. By doing this, the system could support initiatives to preserve ecosystem balance in addition to increasing productivity.

Although the proposed system has the potential to revolutionize modern agriculture, it is important to note that field validation has yet to be conducted. Future research will include comprehensive field trials to assess the system's performance under actual farming conditions. In terms of technical limitations, the current prototype has been designed for use in general outdoor environments but is not fully optimized for harsh field conditions such as heavy rain or extreme weather. In addition, one major challenge is the system's reliance on Wi-Fi or cellular networks for data transmission. In remote rural areas where connectivity is limited or unavailable, the system cannot transmit sensor data in real time, so crop recommendations cannot be determined. Overcoming this limitation will be a major focus of future system upgrades, potentially through the integration of local storage or offline processing capabilities.

ACKNOWLEDGEMENTS

This research activity is supported through Riset dan Inovasi untuk Indonesia Maju (RIIM) Kompetisi funding from the Indonesia Endowment Fund for Education Agency, the Ministry of Finance of the Republic of Indonesia, and the National Research and Innovation Agency of Indonesia (BRIN), according to contract numbers 128/IV/KS/11/2022 and 374/SAM4/PPM/2022. We also thank Telkom University, Indonesia, and Unit Pelaksana Teknis (UPT) Pertanian Wilayah 5 Kabupaten Garut, Indonesia, for supporting this research activity.

REFERENCES

Aditya, B. R., Hernawati, E., Gunawan, T., & Aji, P. (2024). Design an agricultural soil and environment monitoring system based on IoT. *Pertanika Journal of Science and Technology*, *32*(6), 2575–2589. https://doi.org/10.47836/pjst.32.6.09

Akintuyi, O. B. (2024). Adaptive AI in precision agriculture: A review: Investigating the use of self-learning algorithms in optimizing farm operations based on real-time data. *Research Journal of Multidisciplinary Studies*, *7*(2), 16–30. https://doi.org/10.53022/oarjms.2024.7.2.0023

Food and Agriculture Organization of the United Nations. (2021). *The state of food and agriculture 2021: Making agrifood systems more resilient to shocks and stresses*. FAO. https://doi.org/10.4060/cb4476en

Gaikwad, S. V., Vibhute, A. D., Kale, K. V., & Mehrotra, S. C. (2021). An innovative IoT based system for precision farming. *Computers and Electronics in Agriculture*, *187*, 106291. https://doi.org/10.1016/j.compag.2021.106291

García, L., Parra, L., Jimenez, J. M., Lloret, J., & Lorenz, P. (2020). IoT-based smart irrigation systems: An overview on the recent trends on sensors and IoT systems for irrigation in precision agriculture. *Sensors*, *20*(4), 1042. https://doi.org/10.3390/s20041042

Iskandar, A. F. (2021). *Kerabatani: Ubiquitous smart farming platform dilengkapi sistem rekomendasi untuk meningkatkan produktivitas pertanian* [Kerabatani: Ubiquitous smart farming platform equipped with a recommendation system to increase agricultural productivity]. [Unpublished Master's thesis]. Universitas Negeri Malang.

Islam, M. R., Oliullah, K., Kabir, M. M., Alom, M., & Mridha, M. F. (2023). Machine learning enabled IoT system for soil nutrients monitoring and crop recommendation. *Journal of Agriculture and Food Research*, *14*, 100880. https://doi.org/10.1016/j.jafr.2023.100880

Khoeri, I., & Mulyana, D. I. (2021). Implementasi machine learning dengan Decision Tree Algoritma C4.5 dalam penerimaan karyawan baru pada Pt. Gitareksa Dinamika Jakarta [Implementation of machine learning with Decision Tree Algorithm C4.5 in new employee recruitment at PT. Gitareksa Dinamika Jakarta]. *Jurnal Sosial dan Teknologi*, *1*(7), 615–623. https://doi.org/10.59188/jurnalsostech.v1i7.126

Kiruthika, S., & Karthika, D. (2023). IoT-based professional crop recommendation system using a weight-based long-term memory approach. *Measurement: Sensors*, *27*, 100722. https://doi.org/10.1016/j.measen.2023.100722

Lestari, I., Akbar, M., & Intan, B. (2023). Perbandingan algoritma *machine learning* untuk klasifikasi *amenorrhea* [Comparison of machine learning algorithms for amenorrhea classification]. *Journal of Computer and Information Systems Ampera*, *4*(1), 32–43. https://doi.org/10.51519/journalcisa.v4i1.371

Lokhande, D. B., Chavan, T., & Patil, D. P. (2022). Cloud computing information loading and security issues review. *International Journal of Research and Analytical Reviews*, *9*(4), 520–526.

Modi, D., Sutagundar, A. V., Yalavigi, V., & Aravatagimath, A. (2021). Crop recommendation using machine learning algorithm. In *5th International Conference on Information Systems and Computer Networks* (pp. 1–5). IEEE. https://doi.org/10.1109/ISCON52037.2021.9702392

Musanase, C., Vodacek, A., Hanyurwimfura, D., Uwitonze, A., & Kabandana, I. (2023). Data-driven analysis and machine learning-based crop and fertilizer recommendation system for revolutionizing farming practices. *Agriculture*, *13*(11), 2141. https://doi.org/10.3390/agriculture13112141

Priyadharshini, A., Chakraborty, S., Kumar, A., & Pooniwala, O. R. (2021). Intelligent crop recommendation system using machine learning. In *5th International Conference on Computing Methodologies and Communication* (pp. 843-848). IEEE. https://doi.org/10.1109/ICCMC51019.2021.9418375

Qiu, Z., Ma, F., Li, Z., Xu, X., Ge, H., & Du, C. (2021). Estimation of nitrogen nutrition index in rice from UAV RGB images coupled with machine learning algorithms. *Computers and Electronics in Agriculture*, *189*, 106421. https://doi.org/10.1016/j.compag.2021.106421

Rajeswari, S., Suthendran, K., & Rajakumar, K. (2018). A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics. *International Journal of Pure and Applied Mathematics*, *118*(8), 365–370.

Ramson, S. R. J., León-Salas, W. D., Brecheisen, Z., Foster, E. J., Johnston, C. T., Schulze, D. G., Filley, T., Rahimi, R., Soto, M. J. C. V., Bolivar, J. A. L., & Malaga, M. P. (2021). A self-powered, real-time, LoRaWAN IoT-based soil health monitoring system. *IEEE Internet of Things Journal*, *8*(11), 9278–9293. https://doi.org/10.1109/JIOT.2021.3056586

Rao, M. S., Singh, A., Reddy, N. V. S., & Acharya, D. U. (2022). Crop prediction using machine learning. In *Journal of Physics: Conference Series* (Vol. 2161, No. 1, p. 012033). IOP Publishing. https://doi.org/10.1088/1742-6596/2161/1/012033

Ridoy, A. A. I., Siddique, M. A. I., & Joyti, O. (2024). A machine learning-driven crop recommendation system with IoT integration. In *6th International Conference on Electrical Engineering and Information and Communication Technology* (pp. 812–817). IEEE. https://doi.org/10.1109/ICEEICT62016.2024.10534479

Senapaty, M. K., Ray, A., & Padhy, N. (2023). IoT-enabled soil nutrient analysis and crop recommendation model for precision agriculture. *Computers*, *12*(3), 61. https://doi.org/10.3390/computers12030061

Sihombing, P. R., & Arsani, A. M. (2021). Comparison of machine learning methods in classifying poverty in Indonesia in 2018. *Jurnal Teknik Informatika*, *2*(1), 51–56. https://doi.org/10.20884/1.jutif.2021.2.1.52

Singh, P., Kaur, M., & Bajaj, R. (2024). An IoT-enabled crop recommendation system utilizing MQTT for efficient data transmission to AI/ML model. In *7th International Conference on Circuit Power and Computing Technologies* (pp. 315–320). IEEE. https://doi.org/10.1109/ICCPCT61902.2024.10672704

Vemulapalli, S., Sri, M. S., Varshitha, P., Kumar, P. P., & Vinay, T. (2024). An experimental analysis of machine learning techniques for crop recommendation. *Nigerian Journal of Technology*, *43*(2), 301–308. https://doi.org/10.4314/njt.v43i2.13

Wang, X. (2022). Managing land carrying capacity: Key to achieving sustainable production systems for food security. *Land*, *11*(4), 484. https://doi.org/10.3390/land11040484

Yundari, V. S. (2017). *Sistem rekomendasi potensi lahan pertanian menggunakan metode* Simple Additive Weighting *(SAW)* [Agricultural land potential recommendation system using the Simple Additive Weighting (SAW) method] [Master's thesis, Universitas Pendidikan Indonesia]. UPI Repository. http://repository.upi.edu/29867/1/S_KOM_1000161_Title.pdf

Pertanika J. Sci. & Technol. 33 (6): 2855 - 2879 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Enhancing Extractive Text Summarization Using Semantic-Based Similarity Matching

Junjie Yang, Keng Hoon Gan* and Jun Wang

School of Computer Sciences, Universiti Sains Malaysia, 11800 Gelugor, Pulau Pinang, Malaysia

ABSTRACT

Due to the exponential growth of online information, the ability to efficiently extract key content and target information without requiring extensive reading is becoming increasingly important for readers. This paper investigates the construction of neural extractive summarization systems by framing the task as a semantic text matching problem. The proposed approach, named MatchDocSum, aligns the source document with potential summaries within a semantic space, leveraging pretrained language model contextual representations to enhance the understanding of their interconnectedness. The goal is to address the limitations of conventional methods, which often struggle with capturing intricate semantic relationships and producing coherent summaries. Hence, this study proposes an enhanced document summary matching framework to investigate three main aspects that affect the outcome of a good summary: document pruning, text embedding, and similarity matching measure within the framework. MatchDocSum was evaluated on the Cable News Network (CNN)/DailyMail dataset, showing competitive results against several baselines, including LEAD and bidirectional encoder representations from transformers (BERT) for extractive summarization (BERTSUM). The results demonstrate that our approach performs better than baseline models in some aspects, achieving Recall-oriented Under for Gisting Evaluation (ROUGE)-1 scores of 43.50, ROUGE-2 scores of 20.45, and ROUGE-L scores of 40.75.

Keywords: Extractive summarization, natural language processing, semantic similarity, text summarization

ARTICLE INFO

Article history:
Received: 10 January 2025
Accepted: 29 April 2025
Published: 29 October 2025

DOI: https://doi.org/10.47836/pjst.33.6.19

E-mail addresses:
yangjunjie2023@student.usm.my (Junjie Yang)
khgan@usm.my (Keng Hoon Gan)
wangjun@student.usm.my (Jun Wang)
* Corresponding author

INTRODUCTION

Document summarization condenses essential information while preserving key details, tailored to specific document types. News articles focus on “who, what, when, where, why, and how”, research papers emphasize findings and implications, legal documents distill cases and decisions (Zhong et al., 2020), and book summaries

© Universiti Putra Malaysia Press

highlight plots and themes. Summarization enables readers to efficiently process vast amounts of information.

Extractive summarization identifies core information by selecting relevant sentences. Techniques like Named Entity Recognition (NER), syntactic analysis, and algorithms like PageRank (PR) (Pradhan et al., 2013) effectively handle legal texts. Scientific literature uses citation data to highlight findings (Beltagy et al., 2020), while news articles rely on machine learning models like T5 (Raffel et al., 2020) for key sentence extraction.

Graph-based methods like TextRank (Mihalcea & Tarau, 2004) and LexRank (Erkan & Radev, 2004) rank sentences based on positioning and frequency. Recent advancements in deep learning, including BERT (Devlin et al., 2018) and BERTSUM (Liu & Lapata, 2019), improve contextual understanding and summarization by leveraging segment embeddings. T5 adopts a versatile text-to-text approach, excelling across domains.

As textual data grows exponentially, summarization becomes crucial for generating concise, accurate summaries in legal, scientific, and news domains. Semantic matching ensures coherence and relevance, with models like BERT capturing deep contextual relationships for complex documents. These advancements enhance the effectiveness and applicability of summarization techniques across diverse fields.

The task of extractive summarization (Nallapati et al., 2016) within the MatchSum framework faces several limitations. While BERTSUM (Liu & Lapata, 2019) effectively extracts sentence-level information, it struggles to capture sequential relationships within sentences, leading to potential loss of contextual coherence. Additionally, BERT's transformer-based architecture operates in parallel, relying on positional embeddings to approximate word order, which may compromise its ability to understand sentence structure (Vaswani et al., 2017).

BERT has limited performance on long texts due to the quadratic complexity of its self-attention mechanism (Devlin et al., 2018) and its fixed-length input constraints (Beltagy et al., 2020), leading to information loss and inefficient summarization. In addition, the text embeddings generated by BERT and robustly optimized BERT pretraining approach (RoBERTa) have difficulty in capturing subtle relationships and long-distance dependencies in complex documents. In contrast, decoding-enhanced BERT with disentangled attention (DeBERTa), proposed by He et al. (2020), enhances comprehension and generation via untangled attention and outperforms BERT and RoBERTa in modeling long texts and complex semantic relationships.

The MatchSum framework's reliance on cosine similarity to select candidate summaries introduces additional shortcomings. Cosine similarity inadequately captures word order and contextual importance, often leading to summaries that lack semantic richness and coherence (Zhong et al., 2020). Inspired by Zhong et al. (2020), the MatchDocSum architecture was designed to match documents with candidate summaries. The name combines "match", "doc", and "sum" to select the summary that best matches the content of

a document in the semantic space. Unlike BERTSUM-based MatchSum, which relies only on cosine similarity, MatchDocSum uses DeBERTa to encode documents and candidate summaries.

This study explores an improvised method for extractive summarization to address these challenges. It uses the recurrent neural network (RNN)-based SummaRuNNer (Nallapati et al., 2016) model to prune documents and compares its performance with BERTSUM (Liu & Lapata, 2019) on the CNN/DailyMail dataset. DeBERTa is tested as a text encoder in the MatchDocSum framework alongside BERT and RoBERTa. The goal is to identify which model creates the best semantic alignment between documents and candidate summaries. Additionally, it evaluates dot product similarity as an alternative to cosine similarity, seeking a more precise and contextually relevant metric for summary evaluations.

MATERIALS

Text summarization has progressed from traditional statistical methods, like term frequency–inverse document frequency (TF-IDF) and LexRank, which focused on term frequencies and sentence similarity but struggled with long texts and complex semantics, to deep learning models such as convolutional neural networks, RNNs, and sequence-to-sequence (Seq2Seq) models. These modern approaches use encoder-decoder architectures to capture global context in abstractive summarization, but often suffer from errors and irrelevant content.

The advent of pre-trained language models has transformed text summarization. BERTSUM (Liu & Lapata, 2019) enhances extractive summarization by incorporating sentence-level classification into BERT, which improves contextual understanding. MatchSum (Zhong et al., 2020) redefines the task as a text-matching problem and optimizes sentence selection to ensure coherence and relevance, achieving high ROUGE scores. However, challenges related to accuracy, coherence, and computational efficiency still exist. This study refines BERTSUM and MatchSum to improve summarization quality and broaden their applicability across diverse domains.

Extractive Summarization Methods

X. Zhang et al. (2019) presented HiBERT, a hierarchical transformer model built on BERT specifically designed to handle long documents. This model processes text at both the word and sentence levels, enabling it to handle long-range dependencies and summarize extended texts effectively.

BERT-based models have also demonstrated flexibility in various domains. Dutulescu et al. (2022) proposed an unsupervised BERT model tailored for summarizing clinical reports, demonstrating its ability to adapt to diverse text types. In another development, Yao et al. (2018) introduced a method combining BERT with reinforcement learning. This

approach uses reward signals such as readability and informativeness to enhance summary quality, advancing the evaluation of summarization models.

Among prominent models, Liu and Lapata (2019) introduced BERTSUM, an extension of BERT (Devlin et al., 2018) tailored for extractive summarization. BERTSUM enhances BERT's input schema by adding [CLS] and [SEP] tokens for sentence representation and uses interval segment embeddings to differentiate document sentences. The model incorporates summarization-focused layers, including classifiers, inter-sentence Transformer layers, and RNNs, to improve document-level feature capture. On the CNN/DailyMail and New York Times datasets, BERTSUM achieved state-of-the-art ROUGE scores, demonstrating its practical potential for extractive summarization.

Nallapati et al. (2016) introduced SummaRuNNer, an RNN-based framework for extractive summarization framed as sequence classification. Using a bi-directional gated recurrent unit (GRU)-RNN, it processes text hierarchically, capturing fine-grained and document-level dependencies. The model's interpretability stems from its consideration of features like richness, salience, novelty, and position. A novel training mechanism generates approximate extractive labels using a greedy algorithm to optimize ROUGE scores, bypassing reliance on extractive labels. SummaRuNNer performed strongly on CNN/DailyMail and New York Times datasets, achieving high ROUGE metrics, making it a robust tool for extractive summarization.

Hybrid and two-stage summarization approaches have also gained traction. These methods typically involve an extraction phase to select key content, followed by a refinement or compression phase. Early works by Alyguliyev (2009), Galanis et al. (2012), and H. Zhang et al. (2019) extracted important fragments and refined them into coherent summaries. Bae et al. (2019), as well as Chen and Bansal (2018), introduced hybrid frameworks incorporating reinforcement learning to bridge content extraction and rewriting. Liu and Lapata (2019) and Zhong et al. (2020) advanced the "extract-then-compress" methodology, training extractors to identify relevant content and condense it into concise summaries, significantly enhancing summarization quality.

Pre-Trained Models Used in Text Summarization

Extractive summarization selects key sentences from a document to create concise summaries. The introduction of BERT (Devlin et al., 2018) revolutionized this field by capturing semantic relationships and contextual information. As a pre-trained transformer-based model, BERT learns deep bidirectional representations by conditioning on both left and right contexts. When fine-tuned for summarization, it consistently outperforms traditional methods. Figure 1 illustrates the BERT architecture.

RoBERTa enhances BERT's performance by refining its pretraining methods. It removes the Next Sentence Prediction (NSP) task, focusing entirely on Masked Language

Modeling (MLM) to better capture contextual relationships. Trained on a larger dataset (over 160 GB), RoBERTa incorporates extended training times, larger batch sizes, and dynamic masking to prevent overfitting and enhances contextual understanding. These optimizations allow RoBERTa to achieve new benchmarks on tasks like General Language Understanding Evaluation (GLUE) and SquAD, and highlight the significant impact of improved pretraining on generalization. Figure 2 illustrates RoBERTa's architecture.

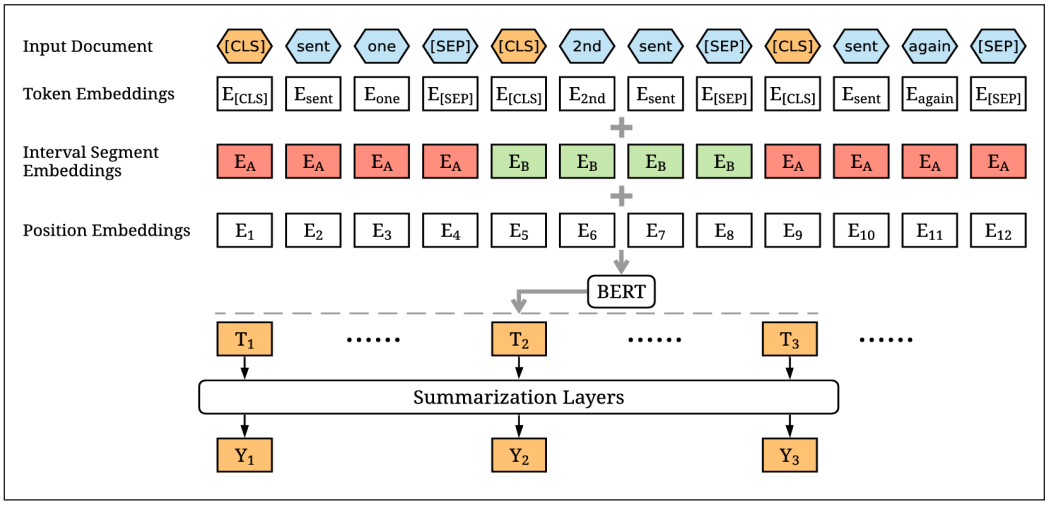


Figure 1. Architecture of bidirectional encoder representations from transformers (BERT) (Liu, 2019)

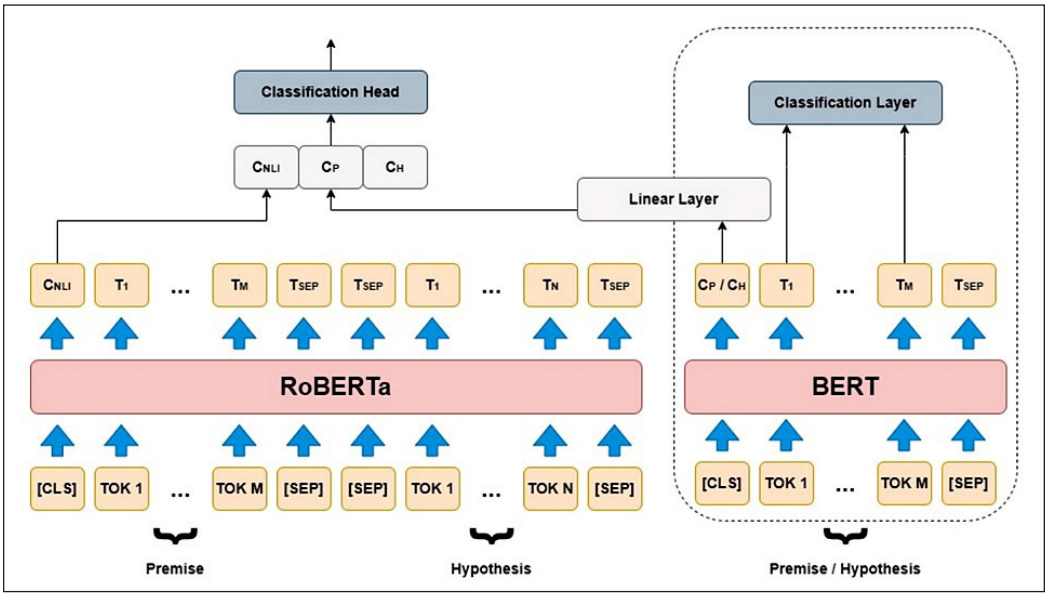


Figure 2. Architecture of robustly optimized BERT pretraining approach (RoBERTa) (Liu et al., 2019)
Note. BERT = Bidirectional encoder representations from transformers

DeBERTa advances BERT by introducing two key innovations: disentangled attention and an enhanced mask decoder. Unlike traditional models, it separates content and position embeddings, which allows it to compute attention scores independently and enhances contextual understanding. The enhanced mask decoder integrates absolute and contextual position embeddings, excelling in tasks requiring precise word order comprehension. These innovations allow DeBERTa to outperform BERT and RoBERTa across benchmarks and demonstrate the effectiveness of disentangled representations and advanced decoding strategies. Figure 3 illustrates DeBERTa's architecture.

Table 1 presents a comparative analysis of various large-scale pre-trained language models on the GLUE benchmark, a widely used suite of natural language understanding (NLU) tasks. As shown in Table 1, DeBERTa has the highest average score of 90.00,

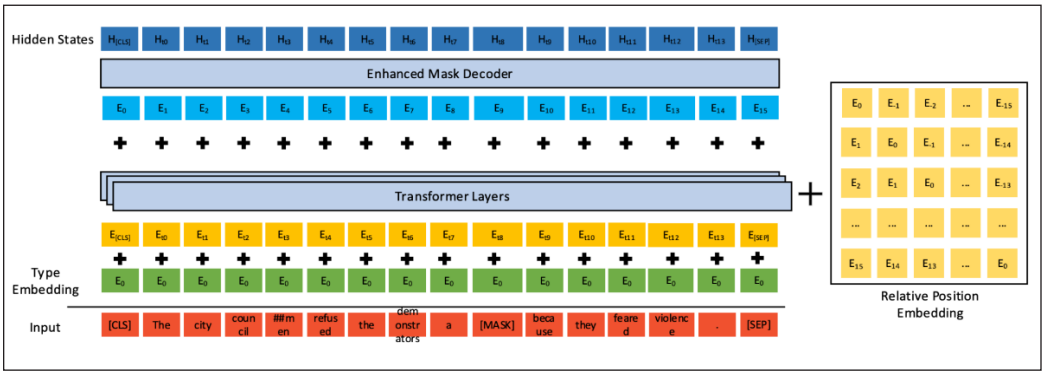


Figure 3. Architecture of decoding-enhanced BERT with disentangled attention (DeBERTa) (He et al., 2021)

Table 1
Comparison results on the General Language Understanding Evaluation (GLUE) development set

Model	CoLA Mcc	OOP Acc	MultiNLI-m/mm Acc	SST/2 Acc	STS-B Corr	QNLI Acc	RTE Acc	MRPC Acc	Avg.
BERT	60.6	91.3	86.6/-	93.2	90.0	92.3	70.4	88.0	84.05
RoBERT	68.0	92.2	90.2/90.2	96.4	92.4	93.9	86.6	90.9	88.82
XLNet	69.0	92.3	90.8/90.8	**97.0**	92.5	94.9	85.9	90.8	89.15
ELECTRA	69.1	**92.4**	90.9/-	96.9	92.6	95.0	88.0	90.8	89.46
DeBERTa	**70.5**	92.3	**91.1/91.1**	96.8	**92.8**	**95.3**	**88.3**	**91.9**	**90.00**

Note. Bolded figures represent the best results; CoLA = Corpus of Linguistic Acceptability; OOP = Out of position; MultiNLI-m/mm = Multi-Genre Natural Language Inference-matched/ mismatched; SST/2 = Stanford Sentiment Treebank (binary version); STS-B = Semantic textual similarity-Benchmark; QNLI = Question Natural Language Inference; RTE = Recognizing textual entailment; MRPC = Microsoft Research Paraphrase Corpus; Mcc = Matthews correlation coefficient; Acc = Accuracy; Corr = Correlation; Avg. = Average; BERT = Bidirectional encoder representations from transformers; RoBERTa = Robustly optimized BERT approach; XLNet = Generalized autoregressive pretraining for language understanding; ELECTRA = Efficiently learning an encoder that classifies token replacements accurately; DeBERTa = Decoding-enhanced BERT with disentangled attention

outperforming the other models. DeBERTa demonstrates its excellent ability to capture linguistic nuances and long-distance dependencies in text.

Background of Document Summary Matching Framework

A Siamese Network, introduced by Bromley et al. (1993), is designed to process and compare two input data points using two identical sub-networks with shared architecture, parameters, and weights. These sub-networks generate feature representations of inputs, mapping them into a latent space where similar points are closer together and dissimilar points are farther apart. After processing, a similarity function (e.g., cosine similarity or dot product) evaluates their relationship. The network uses contrastive loss to minimize distances between similar pairs and maximize distances for dissimilar ones (Koch et al., 2015).

Building upon the Siamese Network framework, Siamese-BERT integrates BERT's contextual embeddings (Devlin et al., 2018) to compute similarity scores between text pairs. Twin BERT models with shared parameters generate embeddings, which are compared using distance metrics to assess semantic similarity. This approach excels in tasks like sentence similarity, paraphrase identification, duplicate question detection, and natural language inference.

The effectiveness of this approach has been extensively validated. Reimers and Gurevych (2019) introduced Sentence-BERT (SBERT), a fine-tuned version of Siamese-BERT, achieving state-of-the-art results on benchmarks such as STS-B and Quora Question Pairs. Applications include identifying duplicate questions on platforms like Quora and Stack Overflow, improving query-document matching in information retrieval, and enhancing reasoning in inference tasks. Siamese-BERT consistently outperforms traditional models like InferSent and Universal Sentence Encoder, benefiting from BERT's rich embeddings and the Siamese network's shared weights, which ensure robust and consistent representations.

A crucial aspect of Siamese-BERT's similarity evaluation is the use of cosine similarity and dot product similarity, which help determine the degree of semantic relatedness between text pairs. Cosine similarity calculates the cosine of the angle between normalized vectors, capturing semantic closeness, while dot product similarity emphasizes vector magnitude, highlighting a sentence's contribution to overall semantic content. These similarity measures are particularly useful in extractive summarization, where they aid in selecting sentences that best preserve a document's core meaning, ensuring coherence and informativeness.

Despite its success, Siamese-BERT faces challenges like high computational costs for fine-tuning and sensitivity to training data quality. Future improvements may involve efficient training techniques like knowledge distillation and incorporating external

knowledge bases or multi-modal data. These advancements could further enhance its performance and applicability.

Research Gap

Despite significant advancements in extractive summarization techniques, critical challenges persist. One key issue is the limited ability of current models to capture nuanced semantic similarities between sentences. While BERT and its variants have improved contextual understanding, they often struggle with subtle semantic distinctions and relationships essential for high-quality summaries.

Most extractive models rely on surface-level features like sentence position or length to rank and select sentences. Although useful, these heuristics fail to capture deeper, more abstract relationships between sentences, resulting in less coherent and contextually accurate summaries. This limitation is especially pronounced in domains like legal or scientific texts, where precise terminology and logical structure are crucial.

Handling long documents remains another challenge. Models like BERTSUM and RoBERTa struggle to maintain coherence in lengthy texts due to fixed input lengths, leading to truncated or incomplete summaries. Although Longformer (Beltagy et al., 2020) provides some improvements for handling longer contexts, it still needs further refinement to preserve the overall semantic flow in complex documents.

Additionally, existing similarity measures, such as cosine similarity, used in frameworks like MatchSum (Zhong et al., 2020), inadequately capture word order and contextual significance. This leads to summaries that fail to fully reflect the intricate relationships in the source text, particularly when word meaning is context-dependent.

In conclusion, current models exhibit limitations in assessing semantic similarity. While methods such as cosine and dot-product similarity are practical, more advanced metrics are needed to preserve semantic integrity. Future efforts should focus on enhancing semantic understanding to produce more coherent, relevant, and semantically rich summaries.

METHODS

This chapter outlines the methodology for developing an extractive text summarization framework using deep learning and the DeBERTa model. By framing extractive summarization as a semantic text alignment problem, the framework improves summary selection through steps including data preprocessing, candidate pruning with SummaRuNNer, and summary generation using DeBERTa. A Siamese-DeBERTa architecture ensures strong semantic alignment between the final summary and the original document, enhancing coherence and relevance.

Formula and Task Definitions

Article (Document) definition: A document D is defined as a sequence of sentences:

$$D = \{s_1, s_2, \ldots, s_n\}$$

where s_i represents the i-th sentence in the document, and n is the total number of sentences.

Pruned document definition: To focus on the most relevant content, the document D undergoes a pruning process, resulting in a pruned document D'. The pruned document D' contains a subset of the original sentences from D:

$$D' = \{s_1, s_2, \ldots, s_m\}$$

where $m \leq n$, and m represents the indices of the selected sentences after pruning.

Candidate summary definition: A candidate summary C is a subset of sentences selected from the pruned document D', and the collection of all possible candidate summaries is denoted as D':

$$C \in D', C = \{s_1, s_2, \ldots, s_k\}$$

where k is the number of sentences in the candidate summary. The sentences $s_1, s_2, \ldots, s_k$ are selected in order they appear in D'.

For example, if D' contains five sentences $\{s_1, s_2, s_3, s_4, s_5\}$, a possible candidate summary C could be $\{s_1, s_3, s_5\}$, retaining important information while forming a concise representation of the document.

For the task definition, the input is a document D consisting of n sentences. The document D is pruned into a subset D', which is then used to generate candidate summaries. The function of the task is to select the candidate summary C from D' that maximizes the semantic similarity between the original document D and the candidate summary C. The output is the candidate summary $\hat{C}$ with the highest semantic similarity.

Let r_D be the embedding vector representing the entire document D, capturing its semantic information. Let r_C be the embedding vector representing a candidate summary C generated from D', capturing its semantic information. The goal is to find the candidate summary C that maximizes the dot product similarity between the embedding vectors r_D and r_C. Formally, the task can be expressed as:

$$\hat{C} = \arg\max_{C \in D}(r_D \cdot r_C)$$

where C is the set of all possible candidate summaries generated from the pruned document D', and $\cdot$ denotes the dot product operation, which measures the similarity between the embedding vectors of the original document D and the candidate summary C.

The task aims to generate an accurate and informative summary $\hat{C}$ with the highest semantic similarity to the original document D. The process begins by pruning D to produce D', a subset of sentences that focuses on the most relevant content. The pruned document D' is then embedded into a vector $r_{D'}$ using a pre-trained language model (e.g., BERT, RoBERTa, or DeBERTa) to capture its semantic information. Multiple candidate summaries C are generated from D', with each summary being a subset of sentences from D'. Each candidate summary C is embedded into a vector r_C using the same pre-trained model.

To measure how well a candidate summary represents the original document, the dot product similarity between r_D (the embedding of D) and r_C (the embedding of each candidate summary) is calculated. The candidate summary with the highest similarity score is selected as the final output. This approach ensures that the selected summary $\hat{C}$ maintains strong semantic alignment with D, resulting in a concise, accurate, and informative summary.

For example, suppose that the pretrained model generates embedding vectors for the source document and candidate summary as r_D = [0.5, 0.7, 0.2] and r_C = [0.4, 0.6, 0.3], respectively. According to the dot product calculation, we first compute the product of each corresponding component ($0.5 \times 0.4 = 0.20$, $0.7 \times 0.6 = 0.42$, $0.2 \times 0.3 = 0.06$) and then sum these values to obtain a dot product similarity of $0.20 + 0.42 + 0.06 = 0.68$. This numerical value intuitively reflects the degree of semantic matching between the candidate summary and the source document in the semantic space, thereby supporting the strategy of selecting the final summary based on the highest similarity score.

Proposed Framework

This section presents the methodology for our enhanced extractive summarization framework, which leverages deep learning techniques and the DeBERTa model with semantic similarity matching to improve text segment selection for summary generation.

Document input: The process begins with an input document containing multiple sentences, which are analyzed to extract the most representative content.

Candidate pruning: Sentences are scored using models like SummaRuNNer, which evaluates content richness, salience, novelty, and positional significance. The top-scoring sentences are selected as candidates for summary generation.

Candidate summary generation: Candidate summaries are created by combining 2–3 top-ranked sentences, ensuring coverage of the document's most important content.

Preprocessing: The document undergoes tokenization, sentence splitting, normalization (e.g., lowercasing and punctuation removal), and truncation to limit length and enhance processing efficiency.

Candidate summary embedding: The Siamese-DeBERTa model generates embedding vectors for both the document and candidate summaries, capturing their semantic relationships. Tied weights ensure consistent encoding, maintaining semantic coherence.

Similarity scoring: Dot product similarity between embedding vectors measures alignment between the document and candidate summaries, capturing both vector magnitude and direction for accurate semantic assessment.

Best candidate summary selection: The candidate summary with the highest similarity score is selected, ensuring alignment with the document's main content and semantic meaning.

Final summary selection: The best candidate summary is refined into the final summary.

This framework integrates deep learning techniques and semantic similarity matching to refine traditional extractive summarization. By framing summarization as a semantic text alignment task, it captures intricate relationships between the document and the summary, producing more coherent and semantically accurate results.

Dot product similarity complements this approach by effectively measuring the alignment between the semantic representations of the document and candidate summaries. Its consideration of vector magnitudes highlights the contribution of each sentence to the document's overall content. When combined with DeBERTa's robust contextual embeddings, dot product similarity enhances the precision of evaluations, yielding more accurate and insightful summaries.

In essence, DeBERTa encodes semantic content, while dot product similarity quantifies the closeness of these representations, ensuring that the final summaries maintain semantic integrity and contextual relevance. Figure 4 illustrates the overall model framework.

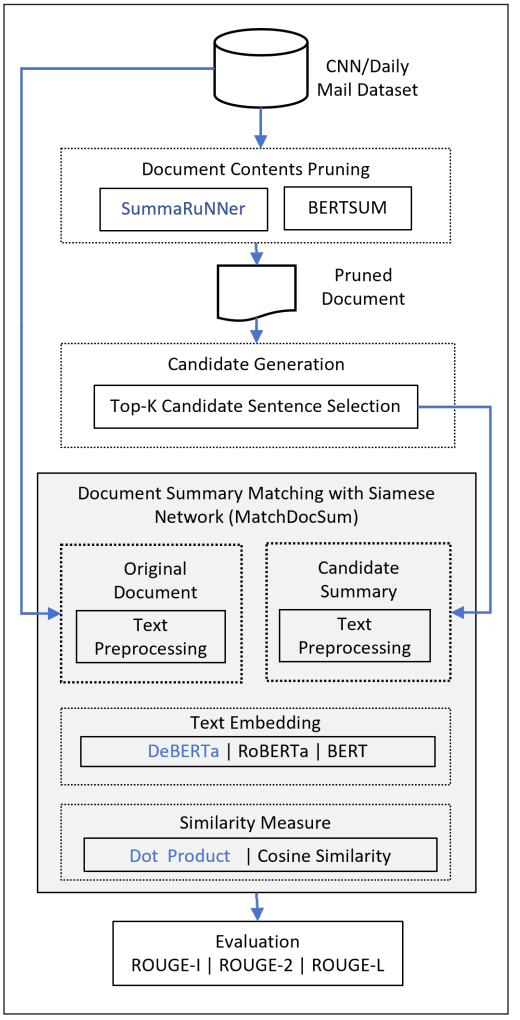


Figure 4. The proposed framework for extractive summarization

Note. CNN = Cable News Network; SummaRuNNer = A recurrent neural network (RNN) based sequence model for extractive summarization of documents; BERTSUM = Fine-tuning BERT for extractive summarization; DeBERTa = Decoding-enhanced BERT with disentangled attention; RoBERTa = Robustly optimized BERT approach; BERT = Bidirectional encoder representations from transformers; ROUGE = Recall-oriented Under for Gisting Evaluation

Datasets

The CNN/DailyMail dataset (Hermann et al., 2015), created for question-and-answer tasks, was later adapted by Nallapati et al. (2016) for text summarization research. Comprising over 300,000 news articles from CNN and Daily Mail, each paired with professionally crafted summaries, the dataset serves as a reliable benchmark for summarization studies. These high-quality summaries are concise yet capture the articles' essential points, enabling rigorous evaluation.

Renowned for its scale, diversity, and quality, the dataset covers topics such as politics, technology, sports, and entertainment. Its extensive training, validation, and testing examples facilitate robust model development and assessment. The CNN/DailyMail dataset remains a cornerstone resource for advancing text summarization and natural language processing research.

Document Contents Pruning

Candidate pruning is the initial step in a two-step summarization process that aims to reduce the document to a smaller subset of sentences, most likely to be included in the final summary. This step optimizes computational efficiency by focusing resources on the most relevant content. The section introduces the candidate-pruning strategy and compares two methods: BERTSUM (Liu & Lapata, 2019), the baseline approach, and SummaRuNNer (Nallapati et al., 2017). By evaluating both approaches, the framework seeks to enhance the effectiveness of sentence selection for summary generation.

A fine-tuned BERTSUM model is used to score sentences based on their probability $P(s_i)$ of being part of the summary, ensuring that the model adapts to the CNN/DailyMail dataset and captures its linguistic and structural patterns more effectively. A threshold τ is applied, either empirically determined or set to select a fixed percentage of top-scoring sentences, to filter out less relevant content. This candidate-pruning method reduces the complexity of the summarization task by narrowing the focus to the most important sentences, thereby optimizing subsequent steps in content generation.

A document $D = \{s_1, s_2, \ldots, s_n\}$ consists of n sentences, each assigned a score $P(s_i)$ that reflects its importance for the summary. These scores are generated using a BERTSUM model fine-tuned for extractive summarization, providing a contextual evaluation of each sentence's relevance. The sentence scoring is defined as:

$$P(s_i) = \text{BERTSUM}(s_i)$$

The probability $P(s_i)$ represents the likelihood that sentence s_i should be included in the summary. Using the BERTSUM model's contextual understanding, scores are computed for each sentence. Candidate sentences are selected by applying a threshold τ, where sentences

with $P(s_i) \geq \tau$ are chosen for further processing. The set of candidates can be represented by the following equation:

$$C = \{ s_i \mid P(s_i) \geq \tau \}$$

In this set C, only the sentences that meet or exceed the threshold t are included, reducing the number of sentences for further analysis. Alternatively, a fixed number of top-scoring sentences can be selected by ranking them in descending order of their scores and choosing the top k sentences:

$$C = \{ s_1, s_2, \ldots, s_k \}$$

where,

$$P(s_1) \geq P(s_2) \geq \ldots \geq P(s_k)$$

Here, C contains the top k sentences with the highest scores, ensuring that the most important sentences are selected.

Another candidate pruning strategy is SummaRuNNer, an RNN-based sequence classifier designed to overcome BERTSUM's limitations in capturing hierarchical structures and sequential dependencies. SummaRuNNer frames extractive summarization as a sequence classification task, dividing documents into sentences, which are further split into words and converted into embeddings.

The model consists of a two-layer bi-directional GRU-RNN. The first layer operates at the word level within each sentence, generating hidden state representations, while the second layer processes these representations at the sentence level to create sentence embeddings. A logistic regression layer assigns binary scores to sentences based on content richness, salience, novelty, and positional importance, determining their inclusion in the summary.

Sentences are ranked by their probabilities of inclusion, and top-ranked sentences are selected as candidates. This approach effectively captures both sentence-level and document-level dependencies, enhancing summary quality.

SummaRuNNer employs GRUs to process sentences, leveraging two gates: the update gate u_j, which retains information from the previous hidden state, and the reset gate r_j, which determines the extent of forgetting prior states (Cho et al., 2014). These gates are computed through specific mathematical operations within the GRU architecture as follows:

$$u_j = \sigma(W_{ux} x_j + W_{uh} h_{j-1} + b_u)$$

$$r_j = \sigma(W_{rx} x_j + W_{rh} h_{j-1} + b_r)$$

The input vector x_j represents the current word, h_{j-1} is the hidden state from the previous time step, W and b are weight matrices and bias vectors, and σ is the sigmoid activation function. The hidden state h_j is updated based on these parameters.

$$h'_j = \tanh(W_{hx}x_j + W_{hh}(r_j \odot h_{j-1}) + b_n)$$

$$h' = (1 - u_j) \odot h'_j + u_j \odot h_{j-1}$$

The final hidden state h_j is a combination of the new hidden state and the previous hidden state h_{j-1}, regulated by the update gate u_j (Cho et al., 2014). For each sentence, a representation is created by concatenating the hidden states from the forward and backward GRUs. The document representation d is then obtained through a non-linear transformation of the average of these concatenated hidden states:

$$d = \tanh\left(W_d\left(\frac{1}{N_d}\sum_{j=1}^{N_d}\left[h_j^f, h_j^b\right]\right) + b\right)$$

The number of sentences in the document is denoted by N_d, and the hidden states of the forward h_j^f and backward h_j^b RNNs are concatenated to form the sentence representation (Nallapati et al., 2017). The probability of a sentence being included in the summary is calculated using a logistic regression model:

$$(y_j = 1 | h_j, s_j, d) = \sigma(W_c h_j + h_j^T W_s d - h_j^T W_r tanh(s_j) + W_a p_a + W_r p_r + b)$$

The terms describe various factors influencing the importance of the j-th sentence in a document for summarization (Nallapati et al., 2017). $W_c h_j$ represents the information content of the j-th sentence. $h_j^T W_s d$ measures its salience concerning the entire document. $-h_j^T W_r \tanh(s_j)$ captures redundancy with respect to the current summary state. $W_a p_a + W_r p_r$ encodes the significance of the absolute and relative positions of the j-th sentence. b is a bias term. The dynamic summary representation s_j, up to the j-th sentence, is calculated as a weighted sum of the hidden states of all previous sentences:

$$s_j = \sum_{i=1}^{j-1} h_i P(y_i = 1 \mid h_i, s_i, d)$$

This formulation integrates sentence-level information, salience, redundancy, and positional importance for summary generation.

Two candidate summary pruning methods were introduced: BERTSUM and SummaRuNNer. BERTSUM leverages a fine-tuned BERT model to score sentences based on embeddings, selecting candidates using a threshold or top-ranked scores. SummaRuNNer utilizes a bidirectional GRU-RNN to generate sentence representations

and scores them through logistic regression, considering content, salience, novelty, and position. By selecting the top *k* sentences, SummaRuNNer effectively captures sequential dependencies and minimizes redundancy. These methods will be compared in experiments to assess their pruning performance.

Candidate Generation

After pruning, the top *k* sentences with the highest scores are selected based on their relevance to the document. Candidate summaries are then generated by forming combinations of 2 or 3 sentences from these top *k* sentences. For example, if $k = 5$, a total of $\binom{5}{2} + \binom{5}{3} = 10 + 10 = 20$ candidate summaries can be created.

Text Preprocessing

Text preprocessing is applied to candidate summaries and the source document to prepare them for analysis. The process includes tokenization, splitting text into smaller units like words or sub words; sentence splitting, converting text to lowercase, and truncation. The maximum sequence length is 512 tokens, including special tokens like [CLS] and [SEP].

Document Summary Matching with Siamese Network

Enhanced semantic matching redefines summarization as a semantic text-matching task, leveraging the DeBERTa model to generate richer embeddings that capture the semantic relationships between documents and candidate summaries. This ensures summaries are coherent and accurately reflect the core content of the original text.

Traditional methods, such as frequency-based approaches (Luhn, 1958) and graph-based methods (LexRank, TextRank), often overlook sentence-level semantic relationships, leading to disjointed, less readable summaries. To overcome this, a modified Siamese-DeBERTa framework is proposed that combines DeBERTa with dot-product similarity to enhance semantic alignment.

Siamese-DeBERTa effectively evaluates semantic similarity between the source document and candidate summaries, preserving the original document's essence while generating high-quality, coherent summaries. This approach improves summary accuracy and relevance, making it a valuable tool for diverse NLP tasks.

Inspired by Zhong et al. (2020), the MatchDocSum architecture aligns a document *D* with a candidate summary *C* in semantic space using DeBERTa. Its Siamese design employs tied weights, ensuring consistent embeddings for both inputs. Representations are derived from the [CLS] markers, and similarity between r_D (document embedding) and r_C (summary embedding) is measured using the dot product, preferred over cosine similarity for capturing vector magnitude and semantic richness.

The MatchDocSum framework utilizes DeBERTa's advanced attention mechanism to capture nuanced semantic relationships, refining text embedding and similarity matching. Pre-trained on a large-scale corpus, DeBERTa produces accurate embeddings, making it highly effective for extractive summarization. Comparative studies confirm that DeBERTa surpasses BERT and RoBERTa in preserving semantic integrity while selecting coherent and relevant sentences.

The framework integrates document pruning, embedding, and matching to deliver high-quality summaries aligned with the source document's content, ensuring both precision and reliability in summary evaluation.

Similarity analysis plays a pivotal role in extractive summarization by ranking candidate summaries based on their semantic similarity to the source document, maintaining coherence and relevance. Common metrics such as cosine and dot product similarity measure semantic closeness, each offering distinct advantages across different applications.

In MatchDocSum, dot-product similarity is selected as the primary matching method due to its ability to capture vector magnitude, enabling more precise semantic evaluation. Research has shown that cosine similarity does not fully capture the relationships in the source text and is particularly limited in representing semantic hierarchy and information weight. By incorporating dot-product similarity, summarization accuracy and quality are enhanced, ensuring that the generated summaries better reflect the semantic richness of the source document. Additionally, cosine similarity is introduced for comparative experiments to further assess the impact of different similarity metrics on summarization effectiveness. Figure 5 illustrates the overall matching process.

A good summary should achieve a higher similarity score than a poor one. To optimize Siamese-DeBERTa, a two-part loss function is employed. The first component, a margin-

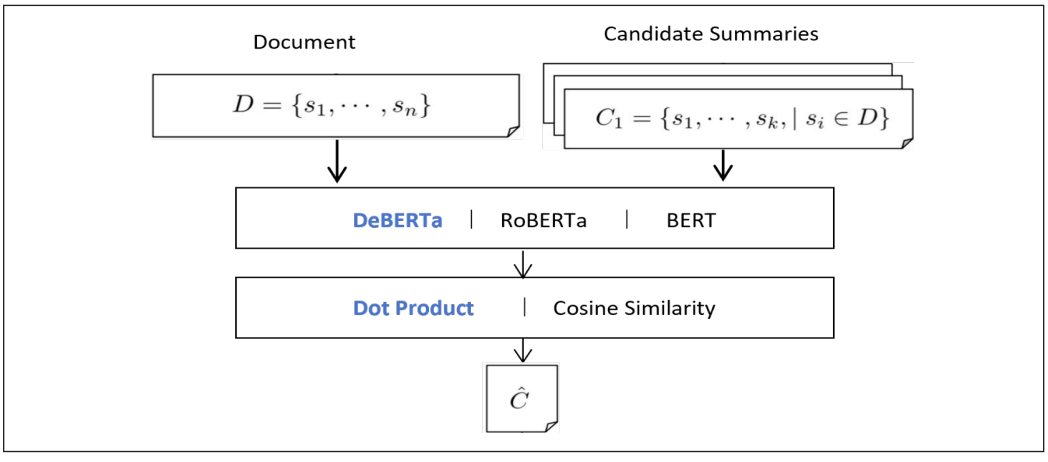


Figure 5. MatchDocSum architecture
Note. DeBERTa = Decoding-enhanced BERT with disentangled attention; RoBERTa = Robustly optimized BERT approach; BERT = Bidirectional encoder representations from transformers

based triplet loss, minimizes the semantic gap between generated and authentic summaries, ensuring fidelity to the source document's core meaning. The second, a ranking gap loss, reinforces score differentiation by minimizing similarity gaps between high- and low-ranking candidates, ensuring high-quality summaries consistently score higher. The loss function is defined as follows:

$$L = \max\left(0, S_C - S_{gt} + \gamma_1\right) + \max\left(0, S_{C_j} - S_{C_i} + \gamma_2 \; (i < j)\right)$$

S_C and S_{gt} represent the similarity values between the candidate summaries, gold summaries, and the source document D, respectively. The i and j denote the order of the candidate summaries, while γ_1 and γ_2 are hyperparameters, with γ_2 being orthogonal to $(i-j)$. By employing this dual loss function, the model considers both semantic proximity to the source document and the ranking gap between candidate summaries, enhancing its ability to produce high-quality, semantically precise summaries.

Evaluation Methods

To evaluate summary quality, ROUGE metrics were employed. Although ROUGE may be limited by abstract diversity constraints, it is widely used in abstracting tasks as a standard assessment metric, is easy to compare, is computationally efficient, and higher scores usually reflect good content coverage. ROUGE-1, ROUGE-2, and ROUGE-L metrics are used to evaluate the quality of generated summaries by comparing them to reference summaries.

ROUGE-1 focuses on the overlap of individual words (unigrams) between the generated summary and the reference summary. It provides a basic measure of how well the hypothesis captures the content of the reference by evaluating word-level similarity.

ROUGE-2 extends this analysis to sequences of words (bigrams). By capturing pairs of consecutive words, it evaluates how well the generated summary preserves the word order and flow of the reference summary. ROUGE-1 and ROUGE-2 calculate the recall of word sequences (unigrams and bigrams), enabling a granular evaluation of text summarization or translation quality.

ROUGE-L measures the longest common subsequence (LCS) between the hypothesis and the reference, allowing for non-contiguous matches. It evaluates similarity at a structural level, combining recall and precision of the LCS to emphasize recall, in alignment with ROUGE's primary objective of measuring content overlap.

RESULTS AND DISCUSSION

Baseline Comparison

This section outlines the baseline models used for comparison with the proposed MatchDocSum model on the CNN/DailyMail dataset:

- LEAD (Nallapati et al., 2016): Selects the first three sentences of each document as the summary, leveraging the inverted pyramid structure of news articles.
- ORACLE (Hirao et al., 2017): Chooses sentences that maximize ROUGE scores relative to reference summaries, representing the upper bound of extractive summarization.
- BERTSUM (Liu & Lapata, 2019): Extends BERT for extractive summarization by introducing interval segment embeddings to capture sentence relationships better.
- MatchSum (Zhong et al., 2020): Frames summarization as a semantic text matching problem, using BERTSUM for sentence encoding and cosine similarity for matching.
- SummaRuNNer (Nallapati et al., 2017): An RNN-based model that selects sentences based on relevance, novelty, and position.

The performance of these models is evaluated using ROUGE metrics for comparison with MatchDocSum.

Experiments on Candidate Pruning

In this experiment, SummaRuNNer was used for candidate pruning in the MatchDocSum model. The two-layer bi-directional GRU-RNN evaluated sentences based on content richness, relevance, novelty, and positional significance. Cosine similarity was employed to select the best sentences, effectively reducing redundancy and improving summary diversity compared to the BERTSUM baseline. The results are summarized in Table 2.

SummaRuNNer combined with RoBERTa achieved the highest ROUGE scores in Table 2: ROUGE-1 of 43.10, ROUGE-2 of 20.10, and ROUGE-L of 40.00, outperforming BERTSUM + BERT, which had the lowest scores (ROUGE-1 of 41.85, ROUGE-2 of 19.34, and ROUGE-L of 39.90). The improvement is attributed to SummaRuNNer's ability to capture sequential dependencies and evaluate multiple features, resulting in summaries that better preserve the document's structure and essential information.

Table 2
Performance comparison of different pruning methods and models

Pruning method	Model	ROUGE-1	ROUGE-2	ROUGE-L
SummaRuNNer	BERT	42.50	19.80	39.50
	RoBERTa	43.10	20.10	40.00
BERTSUM	BERT	41.85	19.34	39.90
	RoBERTa	42.00	19.50	40.10

Note. ROUGE = Recall-oriented Under for Gisting Evaluation; SummaRuNNer = A recurrent neural network (RNN) based sequence model for extractive summarization of documents; BERTSUM = Fine-tuning BERT for extractive summarization; RoBERTa = Robustly optimized BERT approach; BERT = Bidirectional encoder representations from transformers

Experiments on Text Embedding Using DeBERTa

This experiment evaluates the effectiveness of text embeddings generated by DeBERTa, BERT, and RoBERTa within a Siamese network, applied to documents pruned using SummaRuNNer and BERTSUM methods. The focus is on assessing the impact of these embeddings on cosine similarity and their ability to preserve the semantic content of the original text. ROUGE-1, ROUGE-2, and ROUGE-L scores are calculated for each combination of pruning method and embedding model to evaluate summary quality.

Results in Table 3 show that DeBERTa consistently outperformed BERT and RoBERTa across all pruning methods, achieving the highest ROUGE scores (43.20) with SummaRuNNer-pruned documents. This indicates DeBERTa's superior ability to maintain contextual integrity and semantic richness in summaries. While BERT and RoBERTa performed well, their scores were slightly lower, emphasizing DeBERTa's advantage for embedding in text summarization tasks, especially when combined with effective pruning methods like SummaRuNNer.

Table 3
Comparison of pruning methods and embedding models on ROUGE metrics

Pruning method	Embedding model	ROUGE-1	ROUGE-2	ROUGE-L
SummaRuNNer	DeBERTa	43.20	20.30	40.50
	BERT	42.50	19.80	39.90
	RoBERTa	43.10	20.10	40.00
BERTSUM	DeBERTa	42.00	19.50	39.70
	BERT	41.85	19.34	39.50
	RoBERTa	42.10	19.60	39.80

Note. ROUGE = Recall-oriented Under for Gisting Evaluation; SummaRuNNer = A recurrent neural network (RNN) based sequence model for extractive summarization; BERTSUM = Fine-tuning BERT for extractive summarization; DeBERTa = Decoding-enhanced BERT with disentangled attention; RoBERTa = Robustly optimized BERT approach; BERT = Bidirectional encoder representations from transformers

Experiment on Embedding Similarity (Dot Product)

This experiment compares dot product and cosine similarity as semantic similarity measures within a Siamese-DeBERTa architecture. Documents pruned by SummaRuNNer and BERTSUM were embedded using Siamese-DeBERTa, and the similarity between embeddings was calculated using both metrics.

Table 4 shows that the dot product consistently outperformed cosine similarity across all ROUGE metrics. For SummaRuNNer, dot product achieved ROUGE-1, ROUGE-2, and ROUGE-L scores of 43.50, 20.45, and 40.75, surpassing cosine similarity's 43.20, 20.30, and 40.50. Similarly, for BERTSUM, dot product scored 42.60, 19.90, and 39.85, exceeding cosine similarity's 42.30, 19.70, and 39.60.

Table 4
Comparison of dot product and cosine similarity for DeBERTa embeddings under different pruning methods

Pruning method	Similarity measure	ROUGE-1	ROUGE-2	ROUGE-L
SummaRuNNer	Dot product	43.50	20.45	40.75
	Cosine similarity	43.20	20.30	40.50
BERTSUM	Dot product	42.60	19.90	39.85
	Cosine similarity	42.30	19.70	39.60

Note. DeBERTa = Decoding-enhanced BERT with disentangled attention; ROUGE = Recall-oriented Under for Gisting Evaluation; SummaRuNNer = A recurrent neural network (RNN) based sequence model for extractive summarization; BERTSUM = Fine-tuning BERT for extractive summarization

The dot product's ability to capture both vector magnitude and direction enhances semantic alignment, yielding more coherent and accurate summaries. These results underscore its superiority over cosine similarity in extractive summarization tasks using DeBERTa embeddings.

Experimental Results Compared to Baseline

Table 5 compares the proposed MatchDocSum framework with baseline methods. LEAD and ORACLE are widely used baselines, with LEAD selecting the first three sentences and ORACLE maximizing ROUGE scores using abstractive summarization principles. ORACLE generally performs better due to its abstraction-based approach. BERTSUM, the primary baseline in this experiment, demonstrated solid performance but suffered from redundancy issues, consistent with previous studies on the CNN/DailyMail dataset.

Re-implementations of MatchSum with BERTSUM using BERT-base, RoBERTa-base, and DeBERTa-base showed incremental gains, with DeBERTa-base achieving the best results by effectively capturing semantic information and reducing redundancy. In comparison, SummaRuNNer with DeBERTa-base outperformed BERTSUM in content richness, diversity, and redundancy reduction, resulting in higher ROUGE scores.

The MatchDocSum framework, which integrates Siamese-DeBERTa with SummaRuNNer for pruning, further improved semantic preservation and reduced redundancy. Although it did not surpass all baselines in every metric, it demonstrated competitive performance. For instance, MatchDocSum achieved a higher ROUGE-L score (40.75) than BERTSUM (DeBERTa-base) at 39.70, though its ROUGE-1 and ROUGE-2 scores (43.50 and 20.45) were slightly lower than BERTSUM's (42.00 and 19.50). These results reflect MatchDocSum's emphasis on semantic coherence over direct sentence extraction.

While MatchDocSum does not fully outperform ORACLE and MatchSum, its strength lies in semantic alignment optimization. ORACLE sets an upper bound on the theoretical optimum by directly selecting sentences that maximize the ROUGE score. MatchSum

Table 5
Comparison of our framework with various baselines on ROUGE metrics

Model	ROUGE-1	ROUGE-2	ROUGE-L
LEAD	40.43	17.62	36.67
ORACLE	52.59	31.23	48.87
BertSumExt(large) (Liu & Lapata, 2019)	43.85	20.34	39.90
MatchSum (BERT-base) (Zhong et al., 2020)	44.22	20.62	40.38
MatchSum (RoBERTa-base) (Zhong et al., 2020)	44.41	20.86	40.55
SummaRuNNer (DeBERTa-base)	43.20	20.30	40.50
BERTSUM (DeBERTa-base)	42.00	19.50	39.70
MatchDocSum (DeBERTa-base+summarunner_prun)	43.50	20.45	40.75
MatchDocSum (Deberta-base+bertsum_prun)	42.60	19.90	39.85

Note. ROUGE = Recall-oriented Under for Gisting Evaluation; LEAD = Lead baseline; ORACLE = Oracle extractive upper bound; BertSumExt(large) = BERT-based extractive summarization (large version); MatchSum (BERT-base) = Extractive summarization as sentence ranking with BERT-base encoder; MatchSum (RoBERTa-base) = Extractive summarization as sentence ranking with RoBERTa-base encoder; SummaRuNNer (DeBERTa-base) = A recurrent neural network based sequence model for extractive summarization (with DeBERTa-base as encoder); BERTSUM (DeBERTa-base) = Fine-tuned DeBERTa-base for extractive summarization; MatchDocSum (DeBERTa-base+summarunner_prun) = Document matching for summarization (with DeBERTa-base encoder + SummaRuNNer-based pruning); MatchDocSum (DeBERTa-base+bertsum_prun) = Document matching for summarization (with DeBERTa-base encoder + BERTSUM-based pruning)

relies on BERTSUM for sentence-level matching to ensure optimal alignment of candidate summaries to the original document. In contrast, MatchDocSum employs DeBERTa for document encoding, which is better able to capture long-distance dependencies and improves the contextual understanding of summaries by matching the semantic representations of candidate summaries with the source document. Although the pruning step may result in a slightly lower ROUGE score than ORACLE, MatchDocSum performs better in terms of semantic consistency and contextual coherence, provides an extractive summarization method that better meets the actual semantic matching requirements, and brings new optimization ideas to the domain.

Lastly, a noticeable result from the performance between different ROUGE metrics used shows that our proposed models are consistent with the baselines. The scores show that, on average, the models can capture a reasonable portion of the words (when order or context is not considered), but struggle to capture the sequential relationships between words. The overall drop from ROUGE-1 to ROUGE-2 shows the limitation of the models in capturing more complex linguistic structures. Compared to ROUGE-1 and ROUGE-2, ROUGE-L's scores are between the formers, which still show it falls short in fully preserving the structure and the flow of the summary. Despite being able to capture individual words (ROUGE-1), it struggles with capturing more complex structures like bigrams (ROUGE-2)

and maintaining the exact sequence of words (ROUGE-L). Since our approach belongs to the category of extractive summarization, ROUGE measures remain a robust performance metric to capture the basic structure of the summary.

CONCLUSION

The proposed research presents an improvised extractive summarization framework, redefining the task as a semantic text matching problem. It incorporates SummaRuNNer for document pruning to reduce redundancy and improve summary diversity, DeBERTa for generating rich semantic embeddings, and dot product similarity for enhanced semantic alignment. Evaluated on the CNN/DailyMail dataset, the framework demonstrates effectiveness, achieving better semantic preservation and contextual accuracy in summaries, though with slightly lower ROUGE scores than ORACLE due to pruning.

This study holds both theoretical and practical significance. Theoretically, it deepens understanding of how pre-trained models like DeBERTa capture semantic relationships between documents and summaries. The integration of dot product similarity further explores how similarity metrics influence summary quality. Practically, the framework has applications in fields such as news aggregation, information retrieval, and automated report generation, where concise, semantically rich summaries are crucial.

Despite its strengths, the research faces limitations. DeBERTa struggles with long-range dependencies in lengthy texts, potentially missing dispersed key relationships. Its high computational cost limits accessibility for resource-constrained researchers. Evaluation on the CNN/DailyMail dataset raises concerns about generalizability to domains like legal or scientific texts, which have distinct structures. Candidate pruning methods like SummaRuNNer rely on surface-level features such as sentence position and length, which may not accurately reflect sentence relevance. Additionally, ROUGE, as the primary evaluation metric, emphasizes lexical overlap but neglects readability, coherence, and informativeness.

Future improvements could address these limitations by enhancing the framework's capacity to handle long-range dependencies using architectures like memory-augmented networks. Techniques such as model compression or knowledge distillation could lower computational costs, enabling real-time applications. Expanding the framework to diverse domains through domain adaptation and fine-tuning can improve generalizability. Advanced candidate pruning methods, such as reinforcement learning, could reduce biases from surface-level features. Finally, combining ROUGE with semantic metrics like BERTScore could offer a more comprehensive evaluation of summary readability, coherence, and informativeness.

ACKNOWLEDGEMENTS

The authors would like to thank Universiti Sains Malaysia, School of Computer Sciences, for their support in this research work.

REFERENCES

Alyguliyev, R. M. (2009). The two-stage unsupervised approach to multidocument summarization. *Automatic Control and Computer Sciences*, *43*, 276–284. https://doi.org/10.3103/S0146411609050083

Bae, S., Kim, T., Kim, J., & Lee, S.-G. (2019). Summary level training of sentence rewriting for abstractive summarization. In L. Wang, J. C. K. Cheung, G. Carenini, & F. Liu (Eds.), *Proceedings of the 2nd Workshop on New Frontiers in Summarization* (pp. 10–20). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-5402

Beltagy, I., Peters, M. E., & Cohan, A. (2020). *Longformer: The long-document transformer*. arXiv. https://doi.org/10.48550/arXiv.2004.05150

Bromley, J., Bentz, J. W., Bottou, L., Guyon, I., Lecun, Y., Moore, C., Säckinger, E., & Shah, R. (1993). Signature verification using a "Siamese" time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence*, *7*(4), 669–688. https://doi.org/10.1142/S0218001493000339

Chen, Y.-C., & Bansal, M. (2018). Fast abstractive summarization with reinforce-selected sentence rewriting. In I. Gurevych & Y. Miyao (Eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (pp. 675–686). Association for Computational Linguistics. https://doi.org/10.18653/v1/P18-1063

Cho, K., van Merriënboer, B., Gulcehre, Ç., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. In A. Moschitti, B. Pang, & W. Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (pp. 1724–1734). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1179

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of deep bidirectional transformers for language understanding*. arXiv. https://doi.org/10.48550/arXiv.1810.04805

Dutulescu, A.-N., Dascalu, M., & Ruseti, S. (2022). Unsupervised extractive summarization with BERT. In *24th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing* (pp. 158–164). IEEE. https://doi.org/10.1109/SYNASC57785.2022.00032

Erkan, G., & Radev, D. R. (2004). LexRank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, *22*, 457–479. https://doi.org/10.1613/jair.1523

Galanis, D., Lampouras, G., & Androutsopoulos, I. (2012). Extractive multi-document summarization with integer linear programming and support vector regression. In M. Kay & C. Boitet (Eds.), *Proceedings of COLING 2012* (pp. 911–926). The COLING 2012 Organizing Committee.

He, P., Liu, X., Gao, J., & Chen, W. (2020). *DeBERTa: Decoding-enhanced BERT with disentangled attention*. arXiv. https://doi.org/10.48550/arXiv.2006.03654

He, P., Liu, X., Gao, J., & Chen, W. (2021). *Microsoft DeBERTa surpasses human performance on the SuperGLUE benchmark*. Microsoft. https://www.microsoft.com/en-us/research/blog/microsoft-deberta-surpasses-human-performance-on-the-superglue-benchmark/

Hermann, K. M., Kočiský, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M., & Blunsom, P. (2015). *Teaching machines to read and comprehend*. arXiv. https://doi.org/10.48550/arXiv.1506.03340

Hirao, T., Nishino, M., Suzuki, J., & Nagata, M. (2017). Enumeration of extractive ORACLE summaries. In M. Lapata, P. Blunsom, & A. Koller (Eds.), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics* (Vol. 1, pp. 386–396). Association for Computational Linguistics. https://aclanthology.org/E17-1037.pdf

Koch, G., Zemel, R., & Salakhutdinov, R. (2015). *Siamese neural networks for one-shot image recognition*. https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf

Liu, Y. (2019). *Fine-tune BERT for extractive summarization*. arXiv. https://arxiv.org/abs/1903.10318

Liu, Y., & Lapata, M. (2019). Text summarization with pretrained encoders. In K. Inui, J. Jiang, V. Ng, & X. Wan (Eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (pp. 3730–3740). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1387

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A robustly optimized BERT pretraining approach*. arXiv. https://doi.org/10.48550/arXiv.1907.11692

Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of Research and Development*, *2*(2), 159–165. https://doi.org/10.1147/rd.22.0159

Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing* (pp. 404–411). Association for Computational Linguistics.

Nallapati, R., Zhai, F., & Zhou, B. (2017). SummaRuNNer: A recurrent neural network based sequence model for extractive summarization of documents. *Proceedings of the AAAI Conference on Artificial Intelligence*, *31*(1), 3075–3081. https://doi.org/10.1609/aaai.v31i1.10958

Nallapati, R., Zhou, B., dos Santos, C., Gülçehre, Ç., & Xiang, B. (2016). Abstractive text summarization using sequence-to-sequence RNNs and beyond. In S. Riezler & Y. Goldberg (Eds.), *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning* (pp. 280–290). Association for Computational Linguistics. https://doi.org/10.18653/v1/K16-1028

Pradhan, S., Moschitti, A., Xue, N., Ng, H. T., Björkelund, A., Uryupina, O., Zhang, Y., & Zhong, Z. (2013). Towards robust linguistic analysis using OntoNotes. In J. Hockenmaier & S. Riedel (Eds.), *Proceedings of the Seventeenth Conference on Computational Natural Language Learning* (pp. 143–152). Association for Computational Linguistics.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, *21*(1), 5485–5551.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In K. Inui, J. Jiang, V. Ng, & X. Wan (Eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (pp. 3982–3992). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1410

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need*. arXiv. https://doi.org/10.48550/arXiv.1706.03762

Yao, K., Zhang, L., Luo, T., & Wu, Y. (2018). Deep reinforcement learning for extractive document summarization. *Neurocomputing*, *284*, 52–62. https://doi.org/10.1016/j.neucom.2018.01.020

Zhang, H., Cai, J., Xu, J., & Wang, J. (2019). Pretraining-based natural language generation for text summarization. In M. Bansal & A. Villavicencio (Eds.), *Proceedings of the 23rd Conference on Computational Natural Language Learning* (pp. 789–797). Association for Computational Linguistics. https://doi.org/10.18653/v1/K19-1074

Zhang, X., Wei, F., & Zhou, M. (2019). HIBERT: Document level pre-training of hierarchical bidirectional transformers for document summarization. In A. Korhonen, D. Traum, & L. Márquez (Eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 5059–5069). Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1499

Zhong, M., Liu, P., Chen, Y., Wang, D., Qiu, X., & Huang, X. (2020). Extractive summarization as text matching. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 6197–6208). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.552

Pertanika J. Sci. & Technol. 33 (6): 2881 - 2910 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Review Article

Interaction Strategies in Augmented Reality-Situated Visualization for Decision Support: A Systematic Literature Review

Ainnecia Yoag[1,2]*, Masitah Ghazali[1] and Farhan Mohamed[1]

[1]Faculty of Computing, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia
[2]Faculty of Computing and Informatics, Universiti Malaysia Sabah, Labuan International Campus, 87000 Federal Territory of Labuan, Malaysia

ABSTRACT

This systematic literature review (SLR) investigates interaction strategies in augmented reality (AR)-situated visualizations that support decision-making tasks. Despite its growing relevance, a gap remains in understanding how users interact with these systems to perform decision-related tasks. This review aims to examine how various interaction modalities contribute to effective user engagement in decision-making contexts. A total of 23 peer-reviewed studies, published between 2016 and 2024, were analyzed using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure methodological rigor and conceptual alignment. The review focuses on device types, manipulation methods, and interaction modalities. Findings indicate that head-mounted displays (HMDs) are commonly used in immersive settings, while hand-held devices (HHDs) are preferred for their portability and affordability. Touchscreen interaction dominates among HHDs, enabling direct manipulation, while gesture and voice commands are more prominent in HMD-based systems. A conceptual mapping aligns these modalities with manipulation methods such as selection, navigation, and filtering based on Brehmer and Munzner's typology. This review also highlights key challenges, including limited collaborative support, usability concerns, and underrepresentation of domains such as industrial training and public events. This work provides a structured foundation for designing AR-situated visualization systems that better support decision-making across diverse application contexts.

Keywords: Augmented reality, decision-making, situated visualization, user interaction

ARTICLE INFO

Article history:
Received: 22 January 2025
Accepted: 28 July 2025
Published: 29 October 2025

DOI: https://doi.org/10.47836/pjst.33.6.20

E-mail addresses:
ainnecia@graduate.utm.my; ainnecia@ums.edu.my (Ainnecia Yoag)
masitah@utm.my (Masitah Ghazali)
farhan@utm.my (Farhan Mohamed)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

Decision-making is a complex cognitive process integral to human behavior, influencing actions across diverse environments and contexts (Darioshi & Lahav, 2021). It involves selecting an option from a set of alternatives, each assessed against specific criteria to achieve a defined goal. Wang and Ruhe (2007) highlights that effective decision-making requires individuals to be well-informed and possess relevant information for evaluating various options. Research into user decision-making processes and the quality of decisions has been a focal point across multiple fields (Ceschi et al., 2017; Carneiro et al., 2019; Stanitsa et al., 2022). Technological advancements have revolutionized how people access, analyze, and utilize information, reshaping decision-making landscapes and influencing cognitive processes and human behavior (Osiurak et al., 2018; Bonicalzi et al., 2023; Shanmugasundaram & Tamilarasu, 2023). In the context of context-aware interactive systems, especially those delivering real-time or location-specific data, effective decision-making depends not only on access to relevant information but also on how that information is presented and interacted with. The way information is presented and interacted with within a user interface (UI) can significantly influence users' decision-making processes (Starke & Baber, 2018; Peng, 2024). However, the increasing volume of available information has led to the challenge of information overload, which can interfere with decision-making by increasing cognitive load (Roetzel, 2019).

With the rise of emerging technologies such as AR, users now engage with information in immersive and spatially anchored formats that reshape decision behaviors. AR refers to a technology that overlays virtual content onto the real world in real time, aligning digital elements with the user's physical environment to support enhanced perception and interaction (Azuma, 1997). A key application of AR in visualization is through the situated visualization approach, which has become prominent in recent information visualization research (Bressa et al., 2022). Situated visualization integrates virtual data within the physical context of a site, thereby enhancing users' understanding of and interaction with their surroundings (White & Feiner, 2009). The defining feature of this approach is contextual relevance, where data is embedded within the environment to support real-world decision-making (Thomas et al., 2018).

In decision-making scenarios, AR-situated visualizations offer immersive and interactive experiences. By blending virtual objects with real-world environments, AR delivers relevant, contextual information, fostering informed decisions and improving overall decision-making processes (Chen et al., 2024). These benefits have been demonstrated across multiple domains, including agriculture (Zheng & Campbell, 2019), smart home environments (Zheng et al., 2022), shopping environments (Xu et al., 2022; Han et al., 2023), and healthcare (Tanbeer & Sykes, 2024). In AR-situated visualization,

decision-making involves users interacting with spatially embedded information to assess options, compare alternatives, and make informed choices within real-world environments.

Despite its advantages, designing effective AR-situated visualizations for decision-making presents challenges. These include ensuring contextual relevance without overwhelming users with excess information, addressing the varying technical skills and visual capabilities of users (Martins et al., 2023) and creating seamless, intuitive integration into real-world environments (Zhu et al., 2024). Marques et al. (2019) explored the use of AR and situated visualization in decision support systems (DSS), highlighting benefits like improved efficiency, better contextual understanding, and interdisciplinary research opportunities. Martins et al. (2022) refined this framework, focusing on dimensions like space, time, and community, but noted limited empirical validation. Zhu et al. (2024) proposed innovative interaction techniques, including eye-based, hand-based, and spatially aware inputs, to improve usability and engagement in public AR-situated visualization settings. Zheng et al. (2024) further advanced the field by introducing a taxonomy for AR data visualization based on semantic data-reality relationships, demonstrating AR's ability to enhance efficiency, reduce cognitive load, and provide context-aware decision support.

While extensive research has explored the role of AR in decision-making, gaps remain in understanding how AR-situated visualizations can be effectively designed to support these processes. This study addresses this gap by analyzing how AR facilitates decision-making through the presentation of situated information. The study focuses on interaction strategies that enable users to engage effectively with AR-situated visualization data, emphasizing gesture-based controls, touch interfaces, and context-aware interactions.

To ensure a structured and unbiased examination of existing literature, this study adopts an SLR approach. A SLR is a research methodology used to systematically collect, identify, and critically analyze existing studies such as journal articles, conference proceedings, books, and dissertations with the aim of summarizing current knowledge on a specific topic and identifying gaps for future research (Carrera-Rivera et al., 2022). This study examines the interconnection of AR-situated visualization, interaction design, and decision-making processes across various fields. The research aims to provide concrete, applicable findings to enhance AR-situated visualization development and application. The review's core objective is to analyze decision-making task performance using AR-situated visualizations. The study's direction was shaped by specific research questions and their underlying reasoning, detailed in Table 1.

This review is important for advancing the design of AR-situated visualization systems and fostering innovation in decision-making applications for many fields.

Table 1
Research questions and corresponding rationales for conducting the SLR study

No.	Research question	Rationale
RQ1	How does AR-situated visualization support and enhance decision-making processes across various domains?	To uncover the potential of AR-situated visualizations by identifying the application areas and decision categories they impact, and demonstrate how they enhance decision-making across diverse fields
RQ2	What are the predominant AR device technologies examined in research on AR-situated visualization to support decision-making processes?	To highlight the technological advancements driving AR-situated visualization research by identifying the predominant AR devices and their critical roles in enabling effective and efficient decision-making processes
RQ3	How do users interact with AR-situated visualizations to support decision-making processes?	To advance the understanding of user engagement with AR-situated visualizations in facilitating information interpretation and decision-making

Note. SLR = Systematic literature review; RQ = Research question; AR = Augmented reality

MATERIALS AND METHODS

The methodology adheres to the PRISMA framework (Moher et al., 2010) and incorporates an adapted version of the flow diagram from Siean and Vatavu (2021), shown in Figure 1.

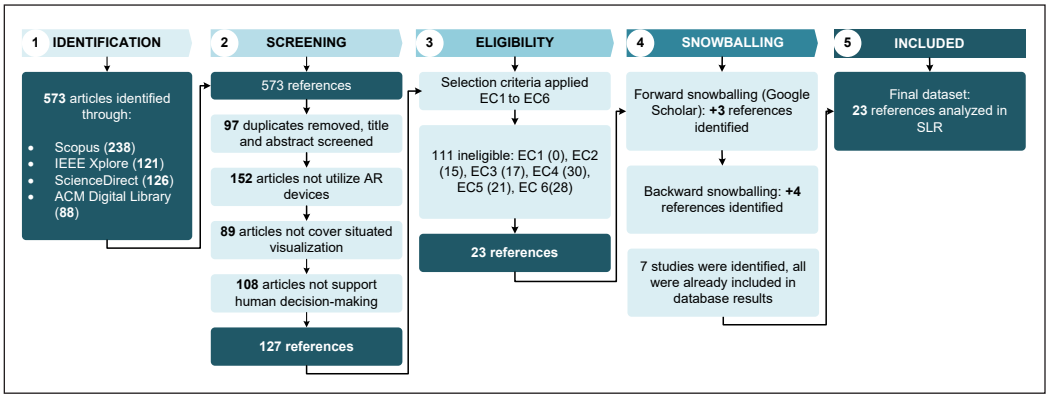


Figure 1. PRISMA-based flow diagram of the study (adapted from Siean & Vatavu, 2021)
Note. PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses; IEEE = Institute of Electrical and Electronics Engineers; ACM = Association for Computing Machinery; AR = Augmented reality; EC = Eligibility criteria; SLR = Systematic literature review

Identification

The SLR employed a meticulous search methodology to locate pertinent research examining the impact of AR-situated visualization on decision-making processes. Multiple prestigious scholarly databases were utilized:

- ScienceDirect,
- Scopus

- ACM Digital Library
- IEEE Xplore

These repositories were chosen due to their comprehensive coverage of AR, human-computer interaction (HCI), and decision-making scholarship.

To ensure both breadth and precision in identifying relevant literature, a staged Boolean search strategy was employed. The initial search queries were constructed using three core concept blocks: (1) AR, (2) situated visualization, and (3) decision-making. This approach aimed to capture a comprehensive set of studies that address the foundational aspects of AR-based DSS. Based on preliminary screening and the review's evolving focus on user interaction, a fourth concept block (4), interaction techniques, was later introduced to refine the search results.

A structured approach was used to craft precise yet inclusive search queries, combining relevant keywords with Boolean operators. Keywords included terms such as "Augmented Reality", "Situated Visualization", and "Decision-Making", along with related phrases like "User Interaction" and "AR Devices". Boolean operators were applied strategically to refine and optimize the search results, with the query structure summarized in Table 2.

Table 2
Boolean search string used to retrieve relevant studies on AR-situated visualization, decision-making, and user interaction for the SLR

Boolean operators
("Augmented Reality" OR "AR") AND ("Situated Visualization" OR "In-situ Visualization OR Visualization") AND ("Decision-Making" OR "Decision Support OR "Decision Task") AND ("User Interaction" OR "Interaction Technique")

Note. AR = Augmented reality; SLR = Systematic literature review

Screening

The database search yielded 573 references, from which 97 duplicate entries were removed, resulting in 476 distinct titles for abstract evaluation. Each abstract underwent thorough analysis to assess its alignment with the research focus on AR-situated visualization and decision-making processes.

The screening procedure eliminated references according to specific criteria:

- 152 papers were discarded due to their absence of AR technology implementation. Though some examined visualization or decision-making in various settings, they fell outside the review's parameters due to a lack of AR components.
- 89 publications were removed for not incorporating situated visualization elements. These works concentrated mainly on standard visualization techniques or AR applications that did not encompass situated visualization principles.

- 108 articles were excluded because they lacked an investigation into decision-making aspects. Despite featuring AR and situated visualization components, these studies did not address the core research objective of exploring the impact of AR-situated visualization on decision-making processes.

Following this comprehensive screening approach, 127 references qualified as suitable and progressed to the eligibility stage.

Eligibility

In the eligibility stage, a comprehensive review was conducted on the complete content of 127 references that had passed the screening phase. This stage employed six specific eligibility criteria (EC), detailed in Table 3, to further narrow down the selection and maintain alignment with the research goals. These carefully crafted criteria served to identify studies of sufficient quality and relevance. Each publication underwent a thorough assessment against these requirements to determine whether it qualified for inclusion. Articles were removed from consideration if they:

- Had publication dates earlier than 2016,
- Were not published in English,
- Examined technologies excluding AR,
- Omitted situated visualization components,
- Did not include decision-making elements, or
- Did not present a concrete system or application.

Publications that satisfied all six criteria were selected for further analysis, while those failing any criterion were eliminated. This meticulous evaluation process yielded a final

Table 3
Eligibility criteria (EC) for selecting studies included in the SLR, detailing both the inclusion and exclusion conditions

No.	Criterion	Inclusion	Exclusion
EC1	Language	Articles written in English	Articles written in non-English
EC2	Timeline	Published between 2016 and 2024	Published before 2016
EC3	Technology deployment	Utilize AR technology for deployment	Does not utilize AR technology
EC4	Articles related to studies	Focus on situated visualization studies	Focus on non-situated visualization studies
EC5	User tasks support	Studies that support decision-making tasks	Studies that do not support decision-making tasks
EC6	Implementation of the system/application	Present an implemented system, prototype, or application of the proposed AR approach	Theoretical or conceptual studies without implementation

Note. SLR = Systematic literature review; AR = Augmented reality

collection of 16 references that specifically matched the study's intended focus, ensuring the maintenance of relevance and uniformity throughout the review process.

Snowballing

A database search was conducted as the primary method, complemented by snowballing techniques to enhance the completeness of the study selection. The snowballing process involved both backward and forward chaining, identifying four studies through backward chaining and three through forward chaining, based on 83 Google Scholar citations. However, all seven studies were already present in the database results, contributing no additional unique articles. This step confirmed the completeness of the initial database search. The final dataset comprised 23 peer-reviewed publications spanning from 2016 to August 2024. The selection process is illustrated in Figure 1.

Data Extraction and Synthesis

The systematic integration of findings from 23 selected studies involved a comprehensive data extraction and synthesis methodology to address the research questions. The extracted information encompassed:

- Basic study identifiers: authors, year of publication, and work title.
- Publication source and publisher information.
- Categories of AR equipment, including portable devices and wearable displays.
- Types of decision-oriented activities.
- User interaction strategies.
- Study constraints and limitations noted by researchers.

The synthesis methodology incorporated dual analytical approaches:

- Quantitative analysis: Identified patterns in AR technology adoption, fields of application, and interaction methodologies, establishing a broad research context.
- Qualitative analysis: Uncovered common themes and challenges across studies.

The results were organized and displayed through structured tables and visual representations to maintain clarity and alignment with the research objective.

Data analysis and classification aligned with specific research questions and their underlying purposes:

- RQ1: Explored domains of application and decision support categories influenced by AR-situated visualizations.
- RQ2: Analyzed technical implementations and AR device classifications in immersive settings.
- RQ3: Investigated user interaction approaches using the taxonomy of interaction techniques for immersive AR proposed by Hertel et al. (2021). This framework

structured the analysis of interaction techniques, mapping them to the multi-level typology of abstract visualization tasks by Brehmer and Munzner (2013). These tasks included *selecting*, *navigating*, *arranging*, *changing*, *filtering*, and *aggregating*, providing a comprehensive perspective on user engagement.

RESULTS AND DISCUSSION

This section presents the results from the 23 reviewed articles, summarized in Table 4. The articles published between 2016 and August 2024 explore AR-situated visualization as an innovative solution to support and enhance decision-making experiences.

Overview of Application Areas and Decision Categories (RQ1)

The existing research reveals various implementations of AR-situated visualization for facilitating decision-making, as shown in Figure 2. The retail sector emerges as the primary domain, representing 35% of studies, with eight investigations centered on improving the shopping experience. Agricultural applications, healthcare, domestic spaces, and maintenance each contribute two studies to the literature. Several other fields have single-study representations, encompassing transportation safety, street navigation, classroom environments, business operations, entomology, sports, and aquaculture.

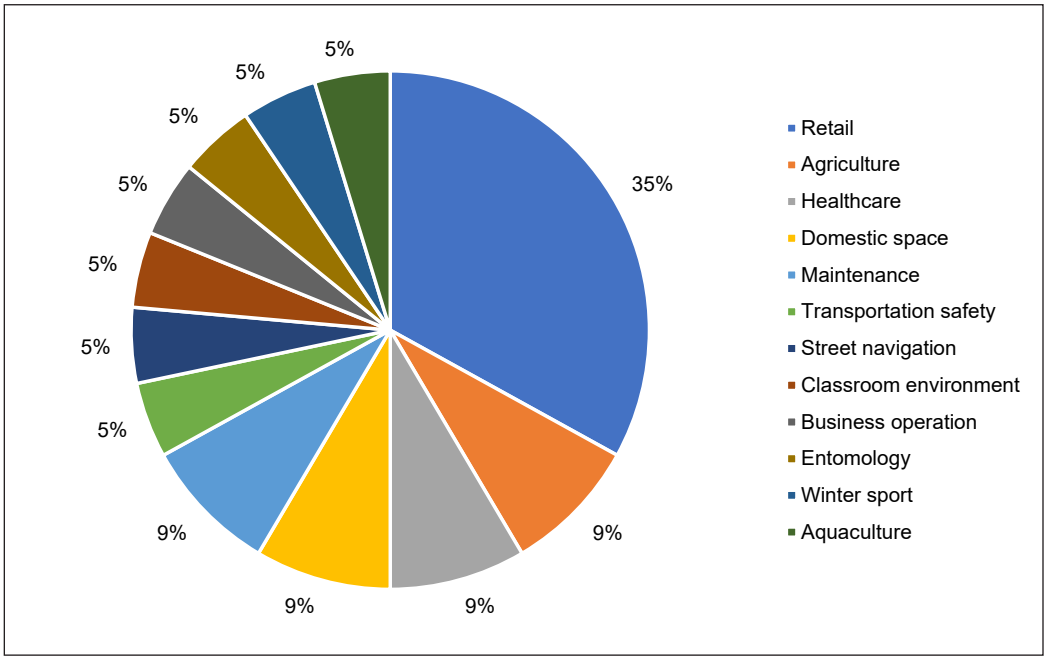


Figure 2. Distribution of application areas identified in studies utilizing augmented reality-based situated visualization from 2016 to August 2024

Table 4
Overview of the studies included in the SLR, summarizing key information such as article title, authorship, publication year, source, publisher, and AR application

No.	Title	Reference	Source	Publisher/Database	AR application/system
1	MiVitals-*Mi*xed Reality Interface for *Vitals* Monitoring: A HoloLens based prototype for healthcare practices	Tanbeer and Sykes (2024)	Computational and Structural Biotechnology Journal	Elsevier/ ScienceDirect	MiVitals
2	BerryScope: AR strawberry picking aid	Tamura et al. (2023)	Proceedings of the 2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct	IEEE	BerryScope
3	PMM: A smart shopping guider based on mobile AR	Han et al. (2023)	Proceedings of the 16th International Symposium on Visual Information Communication and Interaction	ACM	PMM
4	BeeAR: Augmented reality beeline navigation for spatial knowledge acquisition	Mazurkiewicz et al. (2023)	Proceedings of the ACM on Human-Computer Interaction	ACM	BeeAR
5	Using smart glasses to support situated decision making in aquaculture	Xi et al. (2023)	Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems	ACM	Pondside Visualization System
6	BikeAR: Understanding cyclists' crossing decision-making at uncontrolled intersections using augmented reality	Matviienko et al. (2022)	Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems	ACM	BikeAR
7	STARE: Augmented reality data visualization for explainable decision support in smart environments	Zheng et al. (2022)	IEEE Access	IEEE	STARE
8	ARShopping: In-store shopping decision support through augmented reality and immersive visualization	Xu et al. (2022)	Proceedings of the 2022 IEEE Visualization Conference	IEEE	ARShopping

Table 4 *(continue)*

No.	Title	Reference	Source	Publisher/Database	AR application/system
9	Informed ShoppAR - visualizing privacy information in augmented reality	Kaiser et al. (2022)	Proceedings of Mensch und Computer 2022	ACM	Informed ShoppAR
10	Augmented situated visualization methods towards electromagnetic compatibility testing	Guarese et al. (2021)	Computers and Graphics (Pergamon)	Elsevier	ASV
11	VacuumCleanAR: Augmented reality-based self-explanatory physical artifacts	Ludwig et al. (2020)	Proceeding of the 2020 Conference on Mensch und Computer	ACM	VacuumCleanAR
12	Augmented situated visualization for spatial and context-aware decision-making	Guarese et al. (2020)	Proceedings of the 2020 International Conference on Advanced Visual Interfaces	ACM	ASV
13	Occupational therapy meets design: An augmented reality tool for assistive home modifications	Aoyama and Aflatoony (2021)	Proceedings of the 2nd Australian Computer-Human Interaction Conference	ACM	Home-modification AR
14	Location-based augmented reality in-situ visualization applied for agricultural fieldwork navigation	Zheng and Campbell (2019)	Adjunct Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality	IEEE	Location-based AR in-situ visualization and fieldwork navigation system
15	Augmented-reality-enhanced product comparison in physical retailing	Márquez and Ziegler (2019)	Proceedings of the 2019 Conference on Mensch und Computer	ACM	AR-based shopping assistant
16	Let's get in touch - Decision making about enterprise architecture using 3D visualization in augmented reality	Rehring et al. (2019)	Proceedings of the Annual Hawaii International Conference on System Sciences	IEEE	AR EAM
17	gAR-age: A feedback-enabled blended ecosystem for vehicle health monitoring	Shah et al. (2018)	Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications	ACM	gAR-age

Table 4 *(continue)*

No.	Title	Reference	Source	Publisher/Database	AR application/system
18	PHARA: An augmented reality grocery store assistant	Gutiérrez et al. (2018)	Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services	ACM	PHARA
19	AR Furniture: Integrating augmented reality technology to enhance interior design using marker and markerless tracking	Viyanon et al. (2017)	Proceedings of the 2nd International Conference on Intelligent Information Processing	ACM	AR Furniture
20	MelissAR: Towards augmented visual analytics of honey bee behaviour	Engelke et al. (2016)	Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems	ACM	MelissAR
21	A health industry perspective on augmented reality as a communication tool in elderly fall prevention	Lo Bianco et al. (2016)	A Health Industry Perspective on Augmented Reality as A Communication Tool in Elderly Fall Prevention	ACM	Augment
22	Design and evaluation of a wearable AR system for sharing personalized content on ski resort maps	Fedosov et al. (2016)	Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia	ACM	SkiAR
23	Situated Analytics: Demonstrating immersive analytical tools with augmented reality	ElSayed et al. (2016)	Journal of Visual Languages and Computing	Elsevier/ ScienceDirect	Situated Analytics

Note. SLR = Systematic literature review; AR = Augmented reality; IEEE = Institute of Electrical and Electronics Engineers; ACM = Association for Computing Machinery; CHI = Computer-human interaction; PMM = Product Magic Mirror; STARE = Semantic Augmented Reality; ASV = Augmented Situated Visualization; EAM = Enterprise architecture management; PHARA = Personal health augmented reality assistant

The studies are further organized into specific decision support categories, as summarized in Table 5, illustrating the context in which AR-situated visualization enhances decision-making across various domains.

Purchasing decision support is particularly well explored. As shown in Table 5, eight papers highlight AR-situated visualization's ability to address challenges such as limited access to detailed product information (Márquez & Ziegler, 2019; Xu et al., 2022; Han et al., 2023), limited access to product privacy information (Kaiser et al., 2022), difficulties in visualizing items in real-world settings (Viyanon et al., 2017), and challenges in product comparison (ElSayed et al., 2016; Márquez & Ziegler, 2019; Xu et al., 2022). AR-situated visualization effectively addresses these issues by overlaying product descriptions, attributes, and user reviews directly onto the physical or virtual environment. It helps consumers visualize how items would function in real-world settings, reducing uncertainty and enabling more confident purchasing decisions (Márquez & Ziegler, 2019; Xu et al.,

Table 5
Decision support categories and corresponding application contexts in AR-situated visualization studies, along with the number of studies and references reviewed

Decision support category	Number of studies	Application context	List of studies
Purchasing decision	8	Groceries	ElSayed et al. (2016); Gutiérrez et al. (2018); Han et al. (2023); Xu et al. (2022)
		Electronic devices	Ludwig et al. (2020); Márquez and Ziegler (2019); Kaiser et al. (2022)
		Home furniture	Viyanon et al. (2017)
Monitoring decision	5	Pond aeration	Xi et al. (2023)
		EMC	Guarese et al. (2021)
		Automotive maintenance	Shah et al. (2018)
		Patient monitoring	Tanbeer and Sykes (2024)
		Honeybee monitoring	Engelke et al. (2016)
Route navigation decision	3	Cycling	Matviienko et al. (2022)
		Skiing and snowboarding	Fedosov et al. (2016)
		Street navigation	Mazurkiewicz et al. (2023)
Home improvement decision	3	Home-modification	Lo Bianco et al. (2016); Aoyama and Aflatoony (2021)
		Smart home	Zheng et al. (2022)
Agricultural management decision	2	Strawberry farm	Tamura et al. (2023)
		Crops farm	Zheng and Campbell (2019)
Seating decision	1	Seat-choosing	Guarese et al. (2020)
Management decision	1	EAM	Rehring et al. (2019)

Note. AR = Augmented reality; EMC = Electromagnetic compatibility; EAM = Enterprise architecture management

2022). Additionally, it supports side-by-side comparisons of similar products (ElSayed et al., 2016; Márquez & Ziegler, 2019; Xu et al., 2022).

Five studies examined AR-situated visualization for monitoring, improving real-time data collection, visualization, and actionable insights. The technology integrates data within physical spaces to address scattered information systems and cognitive load. Applications range from holographic patient monitoring (Tanbeer & Sykes, 2024) to vehicle maintenance diagnostics (Shah et al., 2018) and electromagnetic compatibility testing (Guarese et al., 2021). Specialized applications include beehive monitoring (Engelke et al., 2016) and aquatic system assessment (Xi et al., 2023). Navigation-focused studies reveal AR's impact on safety enhancement, spatial awareness, and team coordination, featuring applications like AR-enhanced cycling navigation (Matviienko et al., 2022), marker-based routing (Mazurkiewicz et al., 2023), and collaborative navigation systems (Fedosov et al., 2016). Agricultural applications incorporate soil analysis and crop management (Zheng & Campbell, 2019), while precision farming utilizes AR for harvest optimization (Tamura et al., 2023). In home modification situations, AR-situated visualization facilitates occupational therapists in evaluating assistive devices through three-dimensional (3D) modelling in actual settings (Lo Bianco et al., 2016; Aoyama & Aflatoony, 2021). Smart home applications integrate IoT data with physical objects, managing information overload (Zheng et al., 2022). The AR-situated visualization helps users choose better seating by showing environmental information right in the place where they are making those choices (Guarese et al., 2020) and also aids in enterprise architecture management (EAM) by providing easier-to-understand 3D images of information technology (IT) infrastructure (Rehring et al., 2019).

AR-situated visualization demonstrates significant potential across various fields and decision contexts. Technology addresses information accessibility issues, cognitive demands, and spatial integration challenges by providing contextualized decision support. However, scalability remains problematic, with many solutions being confined to specific domains or controlled environments, thereby limiting their broader application. Future developments should emphasize expandable designs and practical implementation studies, focusing on scalability, group interactions, and practical application challenges to maximize the transformative impact of AR.

Predominant Display Device Technologies for AR Situated Visualization in Decision Making (RQ2)

This review categorizes and discusses devices based explicitly on those demonstrated in user testing and evaluation studies. Table 6 presents the specific devices utilized in each research study.

The literature review highlights HMDs and HHDs as the primary input and output devices to deploy AR applications and prototypes that assess AR-situated visualization

Table 6
Predominant device technologies used in AR-situated visualization to support the decision-making process, categorized by input/output device types, number of studies, and cited sources

Device technology	Input/output device	Numbers	Studies
HMD	MR Headset	10	Márquez and Ziegler (2019); Rehring et al. (2019); Guarese et al. (2020, 2021); Ludwig et al. (2020); Matviienko et al. (2022); Zheng et al. (2022); Mazurkiewicz et al. (2023); Tamura et al. (2023); Tanbeer and Sykes (2024)
	Smart glasses	1	Xi et al. (2023)
	VR Headset	1	Zheng and Campbell (2019)
	Smartphone-based HMD (head-worn mount)	1	Fedosov et al. (2016)
HHD	Smartphone	5	Viyanon et al. (2017); Gutiérrez et al. (2018); Shah et al. (2018); Kaiser et al. (2022); Xu et al. (2022)
	Tablet	7	ElSayed et al. (2016); Engelke et al. (2016); Lo Bianco et al. (2016); Viyanon et al. (2017); Zheng and Campbell (2019); Aoyama and Aflatoony (2021); Han et al. (2023)
Wearable devices	Smartwatch (input only)	1	Fedosov et al. (2016)

Note. AR = Augmented reality; HMD = Head-mounted displays; HHD = Hand-held devices; MR = Mixed reality; VR = Virtual reality

approaches in decision-making support. HMDs are devices that are worn on the head and allow virtual content to be displayed directly in the user's field of view (FOV). In contrast, HHDs are portable mobile devices operated manually by users, employing video-see-through technology through rear-facing cameras to seamlessly integrate virtual content with the real-world environment in AR applications (Leins et al., 2024).

HMDs were employed in 13 of the reviewed studies, as shown in Table 6. Among these, MR headsets, particularly the Microsoft HoloLens, were the most frequently utilized, appearing in 10 out of 23 studies, as shown in Table 6. These headsets are favored for their advanced spatial tracking features and mapping capabilities (Guarese et al., 2020, 2021; Mazurkiewicz et al., 2023; Tamura et al., 2023), their hands-free operation with natural interaction that allows users to maintain a head-up experience (Rehring et al., 2019; Zheng et al., 2022; Mazurkiewicz et al., 2023), and their high-quality visualization and display performance (Mazurkiewicz et al., 2023; Tanbeer & Sykes, 2024). In addition to mixed reality (MR) headsets, smart glasses were employed by Xi et al. (2023) in pond water management, highlighting practical considerations for aquaculture tasks, including ruggedization support, lightweight design, computing power, camera capability, and developer-friendly operating systems. Zheng and Campbell (2019) adopted a virtual reality

(VR) headset (HTC VIVE Focus), which provides a wide 110-degree FOV beneficial for viewing wide agricultural fields. However, due to the absence of a built-in Global Positioning System (GPS), smartphone GPS was used for location tracking. Lastly, Fedosov et al. (2016) implemented a head-worn AR system using a smartphone mounted in a head-worn holder that acted as a video see-through display, accompanied by a smartwatch serving as the input device.

In contrast to HMDs, HHDs, which include smartphones and tablets, were employed in 11 studies. Gutiérrez et al. (2018) and Xu et al. (2022) emphasized smartphones' widespread availability and user familiarity, positioning them as the most accessible AR platform. Han et al. (2023) noted HMDs' impracticality and expense in consumer applications like AR shopping, reinforcing HHDs such as tablets as more viable alternatives. These studies underscore HHDs' utility and cost-efficiency, particularly where mobility, affordability, and user-friendliness are crucial. HHDs are also particularly valuable in fieldwork and collaborative scenarios due to their portability. Engelke et al. (2016) developed a tablet-based AR system for bee monitoring, facilitating mobile data access for beekeepers and researchers. Similarly, Aoyama and Aflatoony (2021) demonstrated how HHDs such as tablets facilitate collaborative decision-making by allowing users to explore 3D models from multiple angles within their real environment. The personal health augmented reality assistant (PHARA) system's architectural flexibility demonstrates AR systems' adaptability across devices and scalability potential, from HHDs to HMDs (Gutiérrez et al., 2018). Moreover, Aoyama and Aflatoony (2021) observed that the larger screen size of tablets enhances both interaction and visualization capabilities in AR environments.

Notably, some studies employed devices that fall into more than one category of input/output technology. For example, Zheng and Campbell (2019) utilized a VR headset (classified under HMDs) and a tablet (classified under HHDs) for device comparison. Meanwhile, Fedosov et al. (2016) implemented a smartphone-based head-worn device (classified under HMDs) alongside a smartwatch (classified as a wearable device) for user input. Additionally, Viyanon et al. (2017) does not specify which device was used during user testing; the categorization is based on the stated compatibility of their AR application with both a smartphone and a tablet. Although such studies appear under multiple device categories, each was counted only once in the total number of reviewed studies (N = 23), ensuring accurate representation without duplication.

As shown in Figure 3, research trends in AR decision-making from 2016 to 2024 show that the use of HMDs has generally increased since 2019, while HHD usage had significant increases in 2016, 2018, and 2022, suggesting they are good for mobile and cost-effective applications. The Y-axis in the figure indicates the number of studies published and included in the review that employed either HMDs or HHDs in AR-situated visualization research.

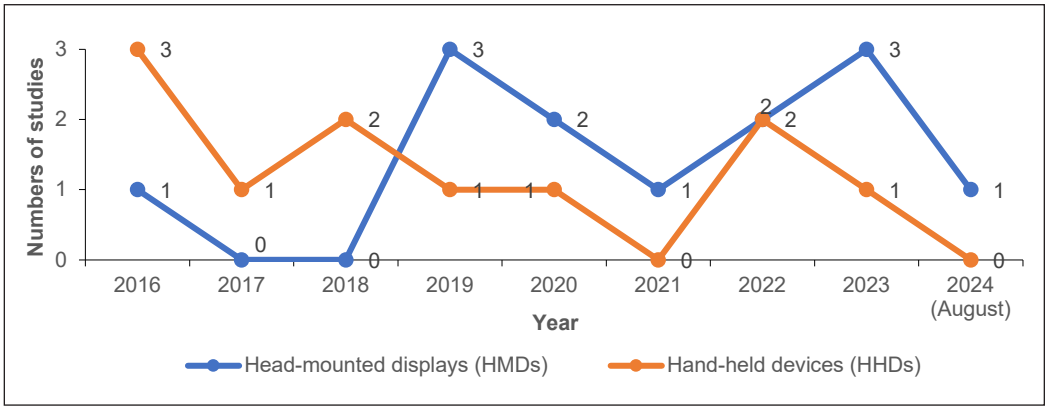


Figure 3. Temporal distribution of augmented reality device usage in reviewed studies from 2016 to August 2024, comparing HMDs and HHDs

The literature review highlights that both HMDs and HHDs offer distinct advantages in AR-situated visualization research. HMDs, such as the Microsoft HoloLens, are particularly effective in scenarios requiring deep immersion, enhanced spatial cognition, and complex interactive features, making them suitable for hands-free applications that demand heightened spatial awareness. Conversely, HHDs like smartphones and tablets provide flexible, cost-effective solutions that are well-suited to dynamic environments with limited resources. Technological advancements have further enhanced the feasibility of HHDs for wider audiences. As noted by Ludwig et al. (2020), while HMDs are advantageous for capturing user attention, smartphones offer greater accessibility. Therefore, the selection between HMDs and HHDs should be guided by the specific constraints and requirements of the decision-making context.

How Do Users Interact with AR Situated Visualization to Support the Decision-Making Process? (RQ3)

This section presents the findings from the literature review on how users interact with AR-situated visualizations during decision-making processes, with a specific focus on HMDs and HHDs. A summary of these findings appears in Table 7, organized using an adapted version of Hertel et al.'s (2021) taxonomy. The analysis categorizes the results based on the input devices, modalities, and interaction methods used, accompanied by corresponding literature citations.

Interaction Through HMDs

Among the 23 studies analyzed, ten employed MR headsets. Six studies (Márquez & Ziegler, 2019; Zheng & Campbell, 2019; Guarese et al., 2020, 2021; Tanbeer & Sykes, 2024) found that users interacted with AR visualizations using hand gestures such as

Table 7
Summary of studies examining how users interact with AR-situated visualization in decision-making contexts, categorized by input device, modality, and interaction techniques, adapted based on the taxonomy of interaction by Hertel et al. (2021)

	Input Device	Modality	Interaction techniques	List of studies
User interaction	MR/VR headset /smart glasses	Gestures	Hand	Márquez and Ziegler (2019); Rehring et al. (2019); Guarese et al. (2020, 2021); Zheng et al. (2022); Tanbeer and Sykes (2024)
		Voice	Voice command	Rehring et al. (2019); Zheng and Campbell (2019); Zheng et al. (2022); Xi et al. (2023); Tanbeer and Sykes (2024)
		Gaze	Head	Márquez and Ziegler (2019); Zheng and Campbell (2019); Ludwig et al. (2020); Guarese et al. (2021)
	Tablet	Tactile interaction	Touch Screen	ElSayed et al. (2016); Engelke et al. (2016); Lo Bianco et al. (2016); Viyanon et al. (2017); Zheng and Campbell (2019); Aoyama and Aflatoony (2021); Han et al. (2023)
	Smartphone			Viyanon et al. (2017); Gutiérrez et al. (2018); Shah et al. (2018); Kaiser et al. (2022); Xu et al. (2022)
	Smartwatch			Fedosov et al. (2016)
	Smart glasses		Touchpad	Xi et al. (2023)
	Clicker device		Clicker	Ludwig et al. (2020)

Note. AR = Augmented reality; MR = Mixed reality; VR = Virtual reality

pointing, tapping on buttons or symbols, and air-tapping (Guarese et al., 2020, 2021; Ludwig et al., 2020). Other ways to interact included zooming, sliding, moving, rotating, panning, and resizing (Rehring et al., 2019; Guarese et al., 2021; Tanbeer & Sykes, 2024). Gesture-based interactions reduced task completion time and supported natural manipulation of in-situ AR visualizations. One notable example is the "touch lens" interaction developed by Guarese et al. (2021), which implements a magnifying glass concept through direct touch. It allows users to access detailed information by simply hovering or tapping over specific data points, which helps minimize hand movement and cognitive load. This interaction is especially beneficial in time-critical decision-making scenarios.

In addition to gesture-based interactions, five studies (Rehring et al., 2019; Zheng & Campbell, 2019; Zheng et al., 2022; Xi et al., 2023; Tanbeer & Sykes, 2024) showed that voice commands can enhance AR interaction by enabling hands-free use. This reduces cognitive load and improves decision-making speed. In applied settings such as healthcare and enterprise management, users relied on voice input to perform tasks like expanding holographic panels, activating visual content, or changing data views. Common voice commands included "show user analysis" and "rotate left" (Rehring et al., 2019), along

with "reset" to return to the original view and "change view" to adjust perspective (Tanbeer & Sykes, 2024). However, real-world environmental challenges such as wind interference and accent variability limited the practical use of voice interaction in Xi et al. (2023), resulting in recognition issues. Despite successful performance in controlled lab conditions, the system failed to interpret spoken commands reliably in the field, leading all participants to abandon voice input in favor of the more consistent touchpad-based tactile interaction.

Four studies examined how AR-situated visualization supports decision-making by investigating gaze modality, particularly head gaze, as an interaction technique. Guarese et al. (2021), Ludwig et al. (2020), Márquez and Ziegler (2019), and Zheng and Campbell (2019) demonstrate that head gaze serves as a hands-free alternative to gesture inputs, allowing for natural interaction with virtual elements through head movement. Head gaze minimizes cognitive burden by matching natural human behaviors and improves effectiveness in product comparison, comprehension, real-time data retrieval, and field operations. However, the studies reveal limitations in head gaze precision and efficiency. Hardware restrictions, such as a limited field of view, diminish interaction effectiveness, particularly in extensive area tasks like agricultural visualization, as observed by Zheng and Campbell (2019). Ludwig et al. (2020) introduced a hybrid interaction combining head movement for orientation with a clicker for selection to overcome usability issues with the HoloLens' "air tap" gesture. Initially, participants struggled with coordination and the limited field of view, but gradually adapted by moving around to view virtual content from different angles. This behavior suggests that spatial exploration can help offset hardware limitations and enhance engagement with AR-situated visualizations during decision-making.

Interaction Through HHDs

Out of the 23 studies examined, 11 specifically investigated AR visualization on HHDs, including tablets and smartphones. Mobile applications were the dominant platform for HHD-based AR. However, one study adopted a web-based AR approach to avoid installation requirements, which raised potential challenges related to internet connectivity in physical settings (Xu et al., 2022). In HHD-based AR applications, users interact with AR content via touchscreens, enabling control of virtual objects through actions such as tapping, swiping, rotating, and dragging. Various functions, such as object selection, deselection, and accessing information, were supported through buttons or icons displayed on the interface (ElSayed et al., 2016; Engelke et al., 2016; Xu et al., 2022; Han et al., 2023).

However, more complex interactions like rotation or drag-and-drop were found to be challenging without clear guidance (Aoyama & Aflatoony, 2021). The heavy reliance on tapping as the primary input method limited interaction flexibility, affecting interface

fluidity and the sense of naturalness (Engelke et al., 2016). Moreover, navigation and task performance were hindered by ambiguous or poorly labelled interface components (Viyanon et al., 2017). While tablets offer a larger display, users generally prefer smartphones due to their convenience and ease of handling. Tablets were associated with ergonomic and usability issues during prolonged use (ElSayed et al., 2016). These findings highlight several areas for improvement: diversifying interaction methods, enhancing interface navigation through clearer labelling, and optimizing the user experience for specific device types.

Multimodal Interaction

The literature review highlights the role of multimodal interaction in enhancing AR visualization applications within decision-making contexts. Several studies demonstrated how combining different input modalities can improve flexibility, usability, and adaptability in task execution. For instance, Rehring et al. (2019) and Tanbeer and Sykes (2024) explored gesture and voice commands in HMD-based systems, while Ludwig et al. (2020) combined head gaze with tactile input through a clicker. Similarly, Zheng et al. (2022) developed a system incorporating image analysis and object recognition with voice commands and hand gestures to facilitate IoT data annotation.

Multimodal interaction provides several advantages. Studies reported that users could switch between modalities based on personal preferences or contextual demands (Rehring et al., 2019; Tanbeer & Sykes, 2024). For users with physical limitations, voice commands served as a practical alternative to hand gestures (Rehring et al., 2019). The ability to select appropriate interaction methods according to situational needs contributed to greater system usability (Ludwig et al., 2020; Tanbeer & Sykes, 2024). In tasks requiring simultaneous actions or divided attention, such as in industrial or field-based settings, combining voice commands with gaze-based input proved effective (Zheng & Campbell, 2019; Zheng et al., 2022).

However, despite these benefits, multimodal systems introduced new challenges. Users often faced steep learning curves when mastering gesture and voice command combinations, and device management complexity was frequently noted, particularly when using HoloLens (Zheng & Campbell, 2019; Ludwig et al., 2020). Reported issues included physical discomfort, limited field of view, and the need for frequent repositioning of the device (Rehring et al., 2019). These findings highlight the trade-offs between interaction flexibility and physical ergonomics. Addressing these limitations remains essential for real-world decision-making applications, and future research should prioritize the development of more intuitive and ergonomically optimized AR devices.

While both HMDs and HHDs enable rich interactive AR experiences, the interaction challenges within AR-situated visualization systems persist. Some studies have proposed reducing user input demands altogether by enhancing passive or minimal-interaction

designs. Fedosov et al. (2016), Matviienko et al. (2022), and Mazurkiewicz et al. (2023) explored the use of persistent visual cues and movement-based guidance to support spatial awareness and real-time decision-making in situations with limited interaction. These designs were especially relevant in time-critical environments such as skiing and cycling, where safety and rapid response were prioritized. Visual overlays guided user attention and navigation while reducing cognitive and physical input load. Although such streamlined approaches demonstrated potential for supporting decisions in fast-paced tasks, their applicability to more complex or multi-step decision scenarios remains underexplored.

Mapping Interaction Modalities to Manipulation Methods

This section elaborates on the categorization presented in Table 7 by mapping interaction modalities, techniques, and user actions to manipulation methods within AR-situated visualization systems. The corresponding diagram in Figure 4 illustrates these mappings, highlighting how different interaction strategies facilitate user engagement across various domains. Based on Brehmer and Munzner's (2013) multi-level typology, this study identifies six key manipulation methods: *select*, *navigate*, *arrange*, *change*, *filter*, and *aggregate*. Mapping these methods to interaction modalities is essential not only for understanding how users interact with AR-situated visualizations through touch, gesture, voice, gaze, or clicker input but also for identifying how users operate on existing visualization elements. The analysis reveals that AR-situated visualizations employ a diverse set of interaction strategies to bridge virtual information with physical contexts, ultimately supporting more efficient and context-aware decision-making. The following details each manipulation method identified in the reviewed studies, beginning with selecting.

Select

Selecting, identified in 20 studies, is a fundamental manipulation method in AR-situated visualization that supports decision-making by enabling users to focus on relevant data points or virtual objects. Among the interaction modalities, touch-based input via HHDs is most frequently used, appearing in 11 studies. These interactions typically involve tapping on 3D objects, UI buttons, or navigation elements. Hand gestures, reported in five studies, support selection through ray-guided pointing, air-tapping, or direct hologram engagement (Márquez & Ziegler, 2019; Rehring et al., 2019; Guarese et al., 2020, 2021; Tanbeer & Sykes, 2024). Voice commands, in two studies, allow hands-free selection via verbal prompts such as "Select [item name]" (Rehring et al., 2019; Tanbeer & Sykes, 2024), offering benefits in hands-occupied contexts. Gaze-driven interaction, observed in two studies, enables selection through head movement, aligning virtual focus with user intention (Márquez & Ziegler, 2019; Zheng & Campbell, 2019).

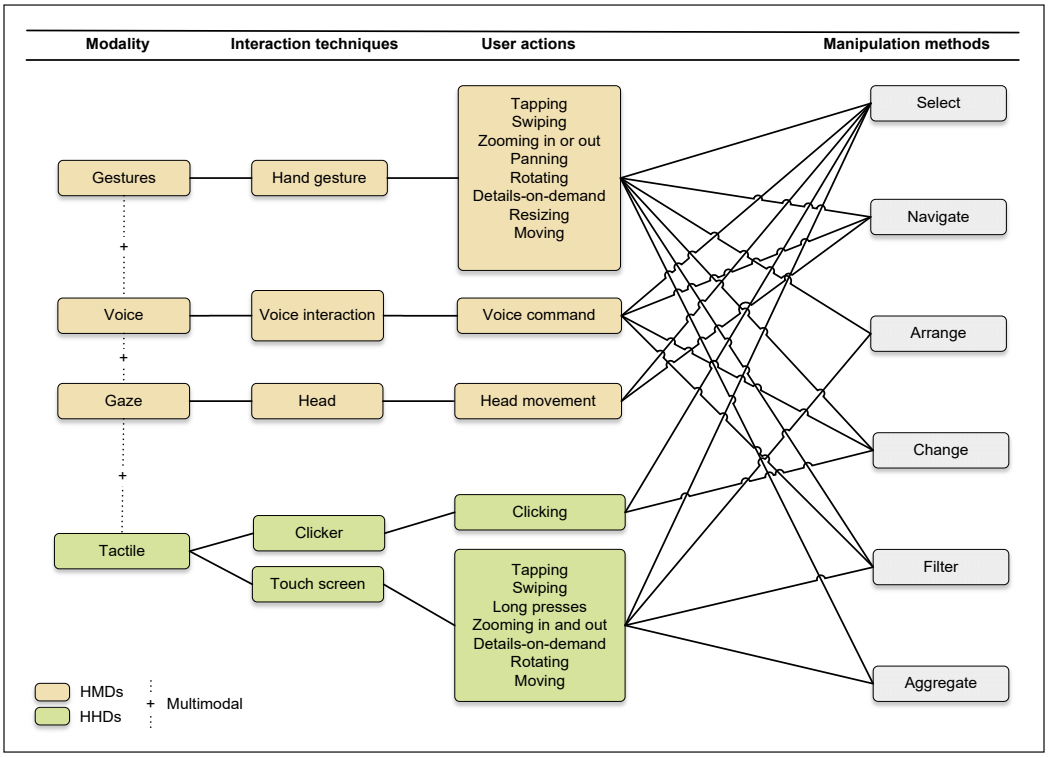


Figure 4. Mapping of interaction modalities and techniques, and user actions in augmented reality (AR)- situated visualization for decision-making tasks, linked to manipulation methods, adapted from Brehmer and Munzner's (2013) typology and Hertel et al. (2021)'s taxonomy of immersive AR interaction techniques
Note. HMDs = Head-mounted displays; HHDs = Hand-held devices

Navigate

Navigating is another core manipulation method used to change the user's viewpoint, allowing them to explore and engage with visual content from multiple perspectives. Supported by 20 studies, navigation enables users to examine objects, access layered information, and interact dynamically with their surroundings. Four studies highlight gestural inputs, such as sliding hands to move panels or zoom, pan, and rotate virtual elements (Rehring et al., 2019; Guarese et al., 2020, 2021; Tanbeer & Sykes, 2024). Three studies describe voice navigation, enabling users to switch screens or perspectives using commands like "go to [screen name]" or "rotate left" (Rehring et al., 2019; Tanbeer & Sykes, 2024), including advanced voice-focus augmentation (Zheng et al., 2022). Head-based navigation, reported in two studies, facilitates orientation and viewpoint shifts through head movement; it can also be used to select options for more information when needed (Márquez & Ziegler, 2019; Ludwig et al., 2020). Touch-based navigation, found in 11 studies, involves tapping icons or buttons to move through environments, switch modes, or manipulate virtual elements such as furniture in 3D spaces.

Arrange

Arranging, found in 11 studies, involves reorganizing visualization elements spatially to optimize layout and improve decision-making accessibility. This method can be performed manually or through system automation. Hand gestures, discussed in two studies, are used to reposition panels or 3D models (Rehring et al., 2019; Tanbeer & Sykes, 2024). Voice commands, as in Rehring et al. (2019) and Tanbeer and Sykes (2024). Arranging visualization assists with commands like "expand blood pressure" to modify panel layout (Tanbeer & Sykes, 2024). Touch screen gestures, also reported in two studies, allow users to hold, drag, rotate, position, or lock objects, which is beneficial for spatial visualization tasks such as configuring assistive technologies in home environments (Viyanon et al., 2017; Aoyama & Aflatoony, 2021). Automated arranging, reported in five studies, dynamically adapts layout based on user attention (Zheng et al., 2022), object orientation (Xu et al., 2022), or feedback (Ludwig et al., 2020; Guarese et al., 2020, 2021), enhancing contextual relevance and ease of access.

Change

Changing refers to modifying the visual encoding or spatial representation of data, identified in 10 studies as a method to adapt content during decision-making. Hand gestures enable resizing and repositioning holographic items or navigating menus in real-time (Rehring et al., 2019; Tanbeer & Sykes, 2024). Voice commands such as "change view" are used to alter perspective or representation (Tanbeer & Sykes, 2024). Touch interactions allow for toggling between data types, updating chart formats (e.g., radar views), or customizing product information (Viyanon et al., 2017; Xu et al., 2022; Han et al., 2023). Gaze-based systems, like those in Márquez & Ziegler (2019), rely on dwell-time activation to trigger visual changes. Automated systems, as in Zheng et al. (2022), highlight or reconfigure content based on focus-tracking. Changes to salient features such as colour and size (ElSayed et al., 2016; Engelke et al., 2016) also support users in comparing or interpreting key data more easily.

Filter

Filtering, reported in 12 studies, helps users narrow visible information by including or excluding elements based on specific attributes. This is particularly useful in selection-based or comparative decision-making tasks. Touchscreen-based filtering, reported in eight studies, allows users to tap buttons to filter by price, content, or features (ElSayed et al., 2016; Engelke et al., 2016; Fedosov et al., 2016; Gutiérrez et al., 2018; Aoyama & Aflatoony, 2021; Kaiser et al., 2022; Xu et al., 2022; Han et al., 2023). Voice commands, used in two studies, include phrases like "open lung details" or "show user analysis" (Rehring et al., 2019; Tanbeer & Sykes, 2024). Other methods involve gaze and gesture

input to manage content visibility, including virtual menus anchored to the user's hand (Márquez & Ziegler, 2019; Guarese et al., 2020). Filtering enables users to focus on relevant data subsets, aiding clarity and enhancing satisfaction.

Aggregate

Aggregating, recognized in three studies, changes the granularity of visualization elements by summarizing or grouping data. For example, touchscreen interactions enable users to calculate total nutritional values by merging data from multiple products (Gutiérrez et al., 2018). AR-based summarization techniques allow collapsing detailed product attributes into broader categories, like "comfort" or "performance," or visualizing generalized overviews of individual elements (Viyanon et al., 2017; Márquez & Ziegler, 2019). This method helps users move between detailed and high-level information, supporting a more efficient understanding of complex data.

The interaction designs in AR-situated visualization enhance users' capacity to manipulate real-time information through context-sensitive and accessible interactions. These manipulation methods form the foundation of how users engage with and adjust existing visualization elements, thereby facilitating informed and situated decision-making. However, some methods (e.g., gestures or gaze) may present usability challenges, especially for non-technical users, reinforcing the importance of intuitive, multimodal interaction designs. Collectively, these manipulation methods contribute to improved accessibility, usability, and engagement in AR applications across domains such as healthcare, retail, and spatial data organization.

DISCUSSION

The findings from RQ1 reveal that AR-situated visualization plays a significant role in enhancing decision-making across diverse domains. In retail, it addresses challenges such as limited product details and comparison by embedding rich contextual information within the physical shopping environment. Beyond retail, applications in healthcare, agriculture, and navigation demonstrate the value of AR-situated visualization in supporting real-time monitoring, spatial understanding, and information clarity. Despite these advantages, several limitations persist, including cognitive overload from excessive information, technical constraints in uncontrolled environments, and ergonomic discomfort during prolonged use. These findings demonstrate the value of designing AR-situated visualization systems that are not only immersive but also adaptable and centered on user needs, thereby improving both decision accuracy and user experience.

The findings for RQ2 indicate that device selection has a critical influence on AR-situated visualization interaction design and its effectiveness in decision-making contexts. HMDs, such as Microsoft HoloLens, enable immersive experiences with features like

gesture and gaze tracking, making them suitable for hands-free, spatially demanding tasks. However, their high cost, physical discomfort, and reduced outdoor usability limit their adoption in broader applications. In contrast, HHDs, including smartphones and tablets, offer greater accessibility, ease of deployment, and suitability for collaborative and mobile contexts, especially in fieldwork or consumer environments. Although they lack the immersive quality of HMDs, their affordability and familiarity make them a practical option. Therefore, device selection should be driven by the interaction needs, task complexity, and environmental constraints of the intended application.

The review for RQ3 lists numerous interaction strategies employed in AR-situated visualizations to support decision-making, each with distinct benefits and challenges. Gesture-based input, commonly used with HMDs, aligns with how users' bodies move more naturally, facilitating intuitive engagement. However, actions like air-tapping may hinder precision for novice users or in tasks requiring fine control. Voice commands offer hands-free operation, but they are sensitive to noise and accent variations, particularly in outdoor or uncontrolled settings. Tactile interaction, especially on touchscreens, offers a reliable alternative for selection and navigation but may limit expressiveness for more complex manipulation.

Consistent with Hertel et al. (2021), this review finds that gesture and voice interactions are dominant in immersive settings. However, unlike previous taxonomies, this study integrates manipulation methods to classify interaction effectiveness for real-time decision-making. Multimodal interfaces that combine gesture, touch, gaze, and voice help balance flexibility and usability, although they may increase learning demands and cognitive load. Across the reviewed studies, selecting, navigating, and filtering emerged as the most frequently supported methods for manipulation. These methods enable users to locate relevant data, explore information spatially, and refine visual content based on specific criteria, directly supporting real-time decision-making. In contrast, less frequently addressed methods, such as arranging, changing, and aggregating, highlight opportunities for further research in supporting more advanced or iterative decision processes. Addressing these underrepresented methods may help expand AR's role in complex decision-support environments.

This study did not include a meta-analysis due to the significant variability in study characteristics, such as application domains, user tasks, metric evaluations, and types of AR devices. Although certain studies demonstrated similarities, the quantity of those with closely aligned focuses, especially regarding outcome measures and experimental design, was insufficient to provide a dependable quantitative synthesis. While previous reviews provided broad overviews of situated visualization (Bressa et al., 2022) or interaction design in AR-situated visualization (Zhu et al., 2024), they did not emphasize how specific interaction modalities support distinct decision-related manipulation methods. This review offers a more decision-focused categorization, mapping interactions to concrete manipulation strategies that align with decision-making goals.

CONCLUSION

This SLR investigated how interaction strategies in AR-situated visualization support the decision-making process across domains, synthesizing findings from 23 selected studies. This review offers a structured classification of interaction modalities such as gesture, touch, voice, and tactile that directly support real-time decision-making, contributing a more decision-centered perspective to the field. It emphasizes the importance of developing AR-situated visualization systems that are not only immersive but also responsive to user needs and situational contexts. Effectiveness depends on aligning visualization and interaction design with decision complexity and cognitive demands, particularly in domains requiring timely and accurate responses.

Device selection also emerged as a critical factor influencing interaction effectiveness. Rather than favoring one platform universally, choices between head-mounted and HHDs should consider task-specific requirements, interaction precision, user mobility, and environmental conditions. Additionally, the review identifies a gap in support for advanced manipulation methods such as arranging, changing, and aggregating. These underexplored areas present valuable directions for future research, especially in complex, collaborative, or iterative decision-making scenarios where current interaction strategies may fall short.

The scope of this review was intentionally narrow to ensure conceptual clarity around decision-making support. As a result, broader AR trends involving exploratory or comparative cognitive activities, which often serve as precursors to decision-making, may not be fully captured. Future reviews would benefit from including these precursor activities and expanding domain coverage to more diverse application areas. The review also identifies areas requiring further exploration, including limited support for advanced or collaborative decision-making and the underrepresentation of domains such as education, public events, and industrial environments. Furthermore, methodological heterogeneity highlights the importance of adopting more standardized research designs to enable broader generalization and cross-study comparability. As a final remark, this review provides a focused foundation for understanding how interaction design in AR-situated visualization can enhance decision support while outlining key directions for future research in increasingly complex and dynamic environments.

ACKNOWLEDGEMENTS

The authors extend their sincere gratitude to the Faculty of Computing, Universiti Teknologi Malaysia (UTM), Johor, Malaysia, and the UTM ViCubeLab for their facilities and research support, which greatly supported this study. The Ministry of Higher Education (MOHE) scholarship for research completion is also acknowledged.

REFERENCES

Aoyama, H., & Aflatoony, L. (2021). Occupational therapy meets design: An augmented reality tool for assistive home modifications. In *Proceedings of the 32nd Australian Conference on Human-Computer Interaction* (pp. 426–436). Association for Computing Machinery. https://doi.org/10.1145/3441000.3441016

Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, *6*(4), 355–385.

Bonicalzi, S., De Caro, M., & Giovanola, B. (2023). Introduction: Digital technologies and human decision-making. *Topoi*, *42*, 793–797. https://doi.org/10.1007/s11245-023-09939-w

Brehmer, M., & Munzner, T. (2013). A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2376–2385. https://doi.org/10.1109/TVCG.2013.124

Bressa, N., Korsgaard, H., Tabard, A., Houben, S., & Vermeulen, J. (2022). What's the situation with situated visualization? A survey and perspectives on situatedness. *IEEE Transactions on Visualization and Computer Graphics*, *28*(1), 107–117. https://doi.org/10.1109/TVCG.2021.3114835

Carneiro, J., Saraiva, P., Conceição, L., Santos, R., Marreiros, G., & Novais, P. (2019). Predicting satisfaction: Perceived decision quality by decision-makers in web-based group decision support systems. *Neurocomputing*, *338*, 399–417. https://doi.org/10.1016/j.neucom.2018.05.126

Carrera-Rivera, A., Ochoa, W., Larrinaga, F., & Lasa, G. (2022). How-to conduct a systematic literature review: A quick guide for computer science research. *MethodsX*, *9*, 101895. https://doi.org/10.1016/j.mex.2022.101895

Ceschi, A., Demerouti, E., Sartori, R., & Weller, J. (2017). Decision-making processes in the workplace: How exhaustion, lack of resources and job demands impair them and affect performance. *Frontiers in Psychology*, *8*, 313. https://doi.org/10.3389/fpsyg.2017.00313

Chen, L., Zhao, H., Shi, C., Wu, Y., Yu, X., Ren, W., Zhang, Z., & Shi, X. (2024). Enhancing multi-modal perception and interaction: An augmented reality visualization system for complex decision making. *Systems*, *12*(1), 7. https://doi.org/10.3390/systems12010007

Darioshi, R., & Lahav, E. (2021). The impact of technology on the human decision-making process. *Human Behavior and Emerging Technologies*, *3*(3), 391–400. https://doi.org/10.1002/hbe2.257

ElSayed, N. A. M., Thomas, B. H., Marriott, K., Piantadosi, J., & Smith, R. T. (2016). Situated Analytics: Demonstrating immersive analytical tools with Augmented Reality. *Journal of Visual Languages and Computing*, *36*, 13–23. https://doi.org/10.1016/j.jvlc.2016.07.006

Engelke, U., Hutson, H., Nguyen, H., & de Souza, P. (2016). MelissAR: Towards augmented visual analytics of honey bee behaviour. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2057–2063). Association for Computing Machinery. https://doi.org/10.1145/2851581.2892367

Fedosov, A., Niforatos, E., Elhart, I., Schneider, T., Anisimov, D., & Langheinrich, M. (2016). Design and evaluation of a wearable AR system for sharing personalized content on ski resort maps. In *Proceedings*

of the 15th International Conference on Mobile and Ubiquitous Multimedia (pp. 141–152). Association for Computing Machinery. https://doi.org/10.1145/3012709.3012721

Guarese, R., Andreasson, P., Nilsson, E., & Maciel, A. (2021). Augmented situated visualization methods towards electromagnetic compatibility testing. *Computers and Graphics*, *94*, 1–10. https://doi.org/10.1016/j.cag.2020.10.001

Guarese, R., Becker, J., Fensterseifer, H., Walter, M., Freitas, C., Nedel, L., & MacIel, A. (2020). Augmented situated visualization for spatial and context-aware decision-making. In *Proceedings of the 2020 International Conference on Advanced Visual Interfaces* (pp. 1–5). Association for Computing Machinery. https://doi.org/10.1145/3399715.3399838

Gutiérrez, F., Verbert, K., & Htun, N. N. (2018). PHARA: An augmented reality grocery store assistant. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (pp. 339–345). Association for Computing Machinery. https://doi.org/10.1145/3236112.3236161

Han, J., Ding, Z., Chen, Y., Yu, L., Chen, S., & Liu, R. (2023). PMM: A smart shopping guider based on mobile AR. In *Proceedings of the 16th International Symposium on Visual Information Communication and Interaction* (pp. 1–8). Association for Computing Machinery. https://doi.org/10.1145/3615522.3615536

Hertel, J., Karaosmanoglu, S., Schmidt, S., Bräker, J., Semmann, M., & Steinicke, F. (2021). A taxonomy of interaction techniques for immersive augmented reality based on an iterative literature review. In *IEEE International Symposium on Mixed and Augmented Reality* (pp. 431–440). IEEE. https://doi.org/10.1109/ISMAR52148.2021.00060

Kaiser, J.-N., Marianski, T., Jung, F., Woźniak, M., & Boll, S. (2022). Informed ShoppAR - Visualizing privacy information in augmented reality. In *Proceedings of Mensch und Computer 2022* (pp. 394–398). Association for Computing Machinery. https://doi.org/10.1145/3543758.3549884

Leins, N., Gonnermann-Müller, J., & Teichmann, M. (2024). Comparing head-mounted and handheld augmented reality for guided assembly. *Journal on Multimodal User Interfaces*, *18*, 313–328. https://doi.org/10.1007/s12193-024-00440-1

Lo Bianco, M., Pedell, S., & Renda, G. (2016). A health industry perspective on augmented reality as a communication tool in elderly fall prevention. In *Proceedings of the International Symposium on Interactive Technology and Ageing Populations* (pp. 1–11). Association for Computing Machinery. https://doi.org/10.1145/2996267.2996268

Ludwig, T., Hoffmann, S., Jasche, F., & Ruhrmann, M. (2020). VacuumCleanAR: Augmented reality-based self-explanatory physical artifacts. In *Proceedings of Mensch und Computer 2020* (pp. 291–302). Association for Computing Machinery. https://doi.org/10.1145/3404983.3405526

Marques, B., Santos, B. S., Araújo, T., Martins, N. C., Alves, J. B., & Dias, P. (2019). Situated visualization in the decision process through augmented reality. In *23rd International Conference Information Visualisation (IV)* (pp. 13–18). IEEE. https://doi.org/10.1109/IV.2019.00012

Márquez, J. O. Á., & Ziegler, J. (2019). Augmented-reality-enhanced product comparison in physical retailing. In *Proceedings of Mensch und Computer 2019* (pp. 55–65). Association for Computing Machinery. https://doi.org/10.1145/3340764.3340800

Martins, N. C., Marques, B., Dias, P., & Santos, B. S. (2022). Augmenting the reality of situated visualization. In *26th International Conference on Information Visualisation (IV)* (pp. 54–60). IEEE. https://doi.org/10.1109/IV56949.2022.00018

Martins, N. C., Marques, B., Dias, P., & Santos, B. S. (2023). Expanding the horizons of situated visualization: The extended SV model. *Big Data and Cognitive Computing*, *7*(2), 112. https://doi.org/10.3390/bdcc7020112

Matviienko, A., Müller, F., Schön, D., Seesemann, P., Günther, S., & Mühlhäuser, M. (2022). BikeAR: Understanding cyclists' crossing decision-making at uncontrolled intersections using augmented reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–15). Association for Computing Machinery. https://doi.org/10.1145/3491102.3517560

Mazurkiewicz, B., de Lima Galvão, M., & Giannopoulos, I. (2023). BeeAR: Augmented reality beeline navigation for spatial knowledge acquisition. In *Proceedings of the ACM on Human-Computer Interaction* (Vol. 7, pp. 1–17). Association for Computing Machinery. https://doi.org/10.1145/3604246

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery*, *8*(5), 336–341. https://doi.org/10.1016/J.IJSU.2010.02.007

Osiurak, F., Navarro, J., & Reynaud, E. (2018). How our cognition shapes and is shaped by technology: A common framework for understanding human tool-use interactions in the past, present, and future. *Frontiers in Psychology*, *9*, 293. https://doi.org/10.3389/fpsyg.2018.00293

Peng, S. (2024). The influence of Graphical User Interfaces on human-computer interaction and the impact of organizing software on decision-making process. *Applied and Computational Engineering*, *50*, 213–221. https://doi.org/10.54254/2755-2721/50/20241509

Rehring, K., Greulich, M., Bredenfeld, L., & Ahlemann, F. (2019). Let's get in touch - Decision making about enterprise architecture using 3D visualization in augmented reality. In *Proceedings of the 52nd Hawaii International Conference on System Sciences* (pp. 1769–1778). University of Hawai'i at Mānoa. https://doi.org/10.24251/hicss.2019.215

Roetzel, P. G. (2019). Information overload in the information age: A review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development. *Business Research*, *12*, 479–522. https://doi.org/10.1007/s40685-018-0069-z

Shah, S., Petluru, S., Singh, R., & Srivastava, S. (2018). gAR-age: A feedback-enabled blended ecosystem for vehicle health monitoring. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 268–277). Association for Computing Machinery. https://doi.org/10.1145/3239060.3239075

Shanmugasundaram, M., & Tamilarasu, A. (2023). The impact of digital technology, social media, and artificial intelligence on cognitive functions: A review. *Frontiers in Cognition*, *2*, 1203077. https://doi.org/10.3389/fcogn.2023.1203077

Sean, A.-I., & Vatavu, R.-D. (2021). Wearable interactions for users with motor impairments: Systematic review, inventory, and research implications. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 1–15). Association for Computing Machinery. https://doi.org/10.1145/3441852.3471212

Stanitsa, A., Hallett, S. H., & Jude, S. (2022). Investigating key factors influencing decision-making in the design of buildings and places: A survey of stakeholders' perception. *Architecture, Structures and Construction*, *2*, 381–401. https://doi.org/10.1007/s44150-022-00058-5

Starke, S. D., & Baber, C. (2018). The effect of four user interface concepts on visual scan pattern similarity and information foraging in a complex decision making task. *Applied Ergonomics*, *70*, 6–17. https://doi.org/10.1016/j.apergo.2018.01.010

Tamura, S., Buayai, P., & Mao, X. (2023). BerryScope: AR strawberry picking aid. In *IEEE International Symposium on Mixed and Augmented Reality Adjunct* (pp. 118–121). IEEE. https://doi.org/10.1109/ISMAR-Adjunct60411.2023.00032

Tanbeer, S. K., & Sykes, E. R. (2024). *MiVitals – Mi*xed Reality Interface for *Vitals* Monitoring: A HoloLens based prototype for healthcare practices. *Computational and Structural Biotechnology Journal*, *24*, 160–175. https://doi.org/10.1016/J.CSBJ.2024.02.024

Thomas, B. H., Welch, G. F., Dragicevic, P., Elmqvist, N., Irani, P., Jansen, Y., Schmalstieg, D., Tabard, A., ElSayed, N. A. M., Smith, R. T., & Willett, W. (2018). Situated analytics. In K. Marriott, F. Schreiber, T. Dwyer, K. Klein, N. H. Riche, T. Itoh, W. Stuerzlinger, & B. H. Thomas (Eds.), *Immersive analytics* (pp. 185–220). Springer. https://doi.org/10.1007/978-3-030-01388-2_7

Viyanon, W., Songsuittipong, T., Piyapaisarn, P., & Sudchid, S. (2017). AR Furniture: Integrating augmented reality technology to enhance interior design using marker and markerless tracking. In *Proceedings of the 2nd International Conference on Intelligent Information Processing* (pp. 1–7). Association for Computing Machinery. https://doi.org/10.1145/3144789.3144825

Wang, Y., & Ruhe, G. (2007). The cognitive process of decision making. *International Journal of Cognitive Informatics and Natural Intelligence*, *1*(2), 73–85. https://doi.org/10.4018/jcini.2007040105

White, S., & Feiner, S. (2009). SiteLens: Situated visualization techniques for urban site visits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1117–1120). Association for Computing Machinery. https://doi.org/10.1145/1518701.1518871

Xi, M., Adcock, M., Thomas, B. H., McCulloch, J., Arnold, S., & George, A. (2023). Using smart glasses to support situated decision making in aquaculture. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–9). Association for Computing Machinery. https://doi.org/10.1145/3544549.3573856

Xu, B., Guo, S., Koh, E., Hoffswell, J., Rossi, R., & Du, F. (2022). ARShopping: In-store shopping decision support through augmented reality and immersive visualization. In *IEEE Visualization and Visual Analytics* (pp. 120–124). IEEE. https://doi.org/10.1109/VIS54862.2022.00033

Zheng, M., & Campbell, A. G. (2019). Location-based augmented reality in-situ visualization 8applied for agricultural fieldwork navigation. In *IEEE International Symposium on Mixed and Augmented Reality* (pp. 93–97). IEEE. https://doi.org/10.1109/ISMAR-ADJUNCT.2019.00039

Zheng, M., Lillis, D., & Campbell, A. G. (2024). Current state of the art and future directions: Augmented reality data visualization to support decision-making. *Visual Informatics*, *8*(2), 80–105. https://doi.org/10.1016/j.visinf.2024.05.001

Zheng, M., Pan, X., Bermeo, N. V., Thomas, R. J., Coyle, D., O'hare, G. M. P., & Campbell, A. G. (2022). STARE: Augmented reality data visualization for explainable decision support in smart environments. *IEEE Access*, *10*, 29543–29557. https://doi.org/10.1109/ACCESS.2022.3156697

Zhu, Q., Wang, Z., Zeng, W., Tong, W., Lin, W., & Ma, X. (2024). Make interaction situated: Designing user acceptable interaction for situated visualization in public environments. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1–21). Association for Computing Machinery. https://doi.org/10.1145/3613904.3642049

REFEREES FOR THE PERTANIKA
JOURNAL OF SCIENCE & TECHNOLOGY

Vol. 33 (6) Oct. 2025

The Editorial Board of the Pertanika Journal of Science and Technology wishes to thank the following:

Abdulrazak Yahya Saleh Al-Hababi
(UNIMAS, Malaysia)

Ahmad Zaidi Abdullah
(UniMAP, Malaysia)

Anshu Bamney
(UConn, USA)

Ariffin Abdul Mutalib
(UUM, Malaysia)

Arun Sharma
(UMTC, USA)

Azmi Murad Masrah Azrifah
(UPM, Malaysia)

C. K. Arvinda Pandian
(BSACIST, India)

Chandan Kumar
(Amrita, India)

Chensong Dong
(Curtin, Australia)

Chua Han Bing
(Curtin, Malaysia)

Denis Jankovic
(UL, Slovenia)

Derya Unlu
(BTU, Turkey)

Diana Carvalho
(UTAD, Portugal)

Dzun Noraini Jimat
(IIUM, Malaysia)

Ebenezer Esenogho
(UNISA, Africa)

Goh Hui Ngo
(MMU, Malaysia)

Haslina Md. Sarkan
(UTM, Malaysia)

Husni Husin
(USK, Indonesia)

Irfana Kabir Ahmad
(UKM, Malaysia)

Juliana Aida Abu Bakar
(UUM, Malaysia)

Kohei Nakanishi
(NU, Japan)

Kumar Pradeep
(CGC, India)

Leong Siew Yoong
(UTAR, Malaysia)

Mahmoud Fathy Mubarak
(EPRI, Egypt)

Mahmud Md Shakir
(VHB, USA)

Mohamed Thariq Hameed Sultan
(UPM, Malaysia)

Mohammad Soleimani Amiri
(UTeM, Malaysia)

Mohd Noh Zarina
(UTeM, Malaysia)

Mohd Shahrul Nizam Mohd Danuri
(UM, Malaysia)

Mostafa Mahmoud El-Gayar
(MANSUniv, Egypt)

Mustafa Ghaderzadeh
(UMSU, Iran)

Nicholas Romano
(VSU, USA)

Nikolay Hinov
(TU-Sofia, Bulgaria)

Nor Hakimin Abdullah
(UMK, Malaysia)

Norhidayah Che Soh
(UMT, Malaysia)

Ramachandran Manikandan
(SASTRA, India)

Sadiq Hussain
(DU, India)

Scheherazed Dakhmouche Djekrif
(UC3, Algeria)

Shahzad Ahmad Qureshi
(PIEAS, Pakistan)

Sofiane Chiheb
(NHSTE, Algeria)

Thirunavukarasu Ramkumar
(VIT, India)

Wang Changling
(CAU, China)

Xie Shaoqu
(GDUT, China)

Yasin Kaya
(ATU, Turkey)

Yusuf Hendrawan
(UB, Indonesia

Amrita	-	*Amrita Vishwa Vidyapeetham*	*UB*	-	*Universitas Brawijaya*
ATU	-	*Adana Alparslan Türkeş Science and Technology University*	*UC3*	-	*University of Constantine 3*
			UConn	-	*University of Connecticut*
BSACIST	-	*B.S. Abdur Rahman Crescent Institute of Science and Technology*	*UKM*	-	*Universiti Kebangsaan Malaysia*
			UL	-	*Univerza v Ljubljani*
BTU	-	*Bursa Teknik Üniversitesi*	*UM*	-	*Universiti Malaya*
CAU	-	*China Agricultural University*	*UMK*	-	*Universiti Malaysia Kelantan*
CGC	-	*Chandigarh Group of Colleges*	*UMSU*	-	*Urmia University of Medical Sciences*
Curtin	-	*Curtin University*	*UMT*	-	*Universiti Malaysia Terengganu*
DU	-	*Dibrugarh University*	*UMTC*	-	*University of Minnesota Twin Cities*
EPRI	-	*Egyptian Petroleum Research Institute*	*UniMAP*	-	*Universiti Malaysia Perlis*
GDUT	-	*Guangdong University of Technology*	*UNIMAS*	-	*Universiti Malaysia Sarawak*
IIUM	-	*International Islamic University Malaysia*	*UNISA*	-	*University of South Africa*
MANSUniv	-	*Mansoura University*	*UPM*	-	*Universiti Putra Malaysia*
MMU	-	*Multimedia University*	*USK*	-	*Universitas Syiah Kuala*
NHSTE	-	*National Higher School of Technology and Engineering*	*UTAD*	-	*Universidade de Trás-os-Montes e Alto Douro*
			UTAR	-	*Universiti Tunku Abdul Rahman*
NU	-	*Nagoya University*	*UTeM*	-	*Universiti Teknikal Malaysia Melaka*
PIEAS	-	*Pakistan Institute of Engineering and Applied Sciences*	*UTM*	-	*Universiti Teknologi Malaysia*
			UUM	-	*Universiti Utara Malaysia*
SASTRA	-	*Shanmugha Arts, Science, Technology & Research Academy*	*VHB*	-	*Vanasse Hangen Brustlin, Inc.*
			VIT	-	*Vellore Institute of Technology*
TU-Sofia	-	*Technical University of Sofia*	*VSU*	-	*Virginia State University*

While every effort has been made to include a complete list of referees for the period stated above, however if any name(s) have been omitted unintentionally or spelt incorrectly, please notify the Chief Executive Editor, *Pertanika* Journals at executive_editor.pertanika@upm.edu.my

Any inclusion or exclusion of name(s) on this page does not commit the *Pertanika* Editorial Office, nor the UPM Press or the university to provide any liability for whatsoever reason.

Performance Evaluation of Jatropha Polyurethane Membrane Modified with Bio-Based Carbon Dots for Pb(II) Removal 2771
Umar Adam Majid, Zurina Zainal Abidin, Nur Haninah Harun and Khairul Faezah Md Yunos

RGB vs. HSV for Kitchen Fire Detection with YOLOv5 2797
Norisza Dalila Ismail, Rizauddin Ramli and Mohd Nizam Ab Rahman

Machine Learning Enabled IoT System for Agricultural Land Recommendation 2833
Bayu Rima Aditya, Anindia Agusta Ken Nadila, Muhammad Qanit Al-Hijran, Muhammad Bintang Ramadhan and Yudha Ginanjar

Enhancing Extractive Text Summarization Using Semantic-Based Similarity Matching 2855
Junjie Yang, Keng Hoon Gan and Jun Wang

Review Article 2881
Interaction Strategies in Augmented Reality-Situated Visualization for Decision Support: A Systematic Literature Review
Ainnecia Yoag, Masitah Ghazali and Farhan Mohamed

Review Article 2565
Exploring Membrane Pervaporation Separation: A Review on Obtaining Combustible Ethanol and Butanol from Fermentation Broths
Jorge Arturo Moreno Cascante, Jairo Antonio Cubillos Lobo and Felipe Bustamante Londoño

Optimisation and Predictive Modelling of Natural Frequencies on Carbon/Glass Hybrid Composite Laminates 2589
Abdul Azim Taredi, Siti Mariam Abdul Rahman, Mohd Nor Azmi Ab Patar and Jamaluddin Mahmud

Case Study 2615
Mid-to-Post-Pandemic Mobility Patterns: A Case Study of Malaysia Highlighting Differential Trends in Work and Non-Work Trips
Megat-Usamah Megat-Johari, Nusayba Megat-Johari, Nischal Gupta, Hisham Jashami and Teik Hua Law

Review Article 2641
Credit Card Fraud Detection Using Machine Learning and Deep Learning Techniques: A Review
Shuchita Sheokand and Sunita Beniwal

Impact of Different Drone Sprayer Nozzle Types on Bio-Fertilizer Application in Wetland Paddy Cultivation 2665
Emmanuel Otache Adanu, Ahmad Suhaizi Mat Su, Azmi Yahya, Siti Amni Ismail, Nik Norasma Che'Ya, Muhammad Huzaifah Mohd Roslim, Abdul Shukor Juraimi, Muhamad Noor Hazwan Abd Manaf and Mst. Motmainna

Review Article 2685
Time Series Prediction for Cycling Sport Performance Through a Machine Learning Predictor
Muhammad Zulazri Hanis Mohd Nawi, Sukhairi Sudin, Fathinul Syahir Ahmad Saad, Muhamad Khairul Ali Hassan, Faranadia Abdul Haris, Nurul Syahirah Khalid and Kamarulzaman Kamarudin

Comparative Study of Nonlinear Regression Models for Cleanliness Factor Prediction in Coal-Fired Utilities 2723
Nur Aina Mohammad Abdul Aziz, Azura Che Soh, Asnor Juraiza Ishak, Nor Mohd Haziq Norsahperi and Amir Redzuan Mohd Ibrahim

Development and Verification of an MCNP5-Based Gamma Camera for ^{177}Lu SPECT Imaging 2747
Puvanesuawary Morthy, Marianie Musarudin, Nor Shazleen Ab Shukor, Mohamad Aminudin Said, Dong Xianling and M. Iqbal Saripan

Pertanika Journal of Science & Technology
Vol. 33 (6) Oct. 2025

Content

Foreword i
Luqman Chuah Abdullah

Considerations of Earth Resistance Reduction Agent Studies 2395
Muhammad Noh Hamzah, Mohd. Zainal Abidin Ab Kadir, Shafreeza Sobri, Chandima Gomes and Syahrun Nizam Md. Arshad Hashim

Review Article 2413
Unveiling Insights on Scientometric Exploration of Trends in Power Cable Research – A Review
Khairunnisa Norfaeza Mohamed Khairul Fariz, Nik Hakimi Nik Ali, Mohd Iqbal Mohd Noor, Dalina Johari, Azrul Mohd Ariffin and Mohd Azraei Pangah @ Pa'at

Review Article 2441
A Comprehensive Review of Tuberculosis Detection and Prediction Using Hybrid Machine Learning Models on Ziehl-Neelsen-Stained Microscopy Images
Nur Amirah Kamaluddin, Murizah Kassim, Nor Salmah Bakar and Shuria Saaidin

Extraction and Characterisation of Natural Fibres from *Pennisetum purpureum*: Influence of Harvesting Age on Chemical, Mechanical, and Thermal Properties 2465
Mohd Fairus Kayat, Ridhwan Jumaidin, Mastura Mohamad Taha and Fahmi Asyadi Md Yusof

Harnessing Co-Culture of Hydrolase–Producing *Bacillus* spp. to Boost Kitchen Waste Biodegradation 2491
Nurhidayah Ramlee, Muhamad Aidilfitri Mohamad Roslan, Helmi Wasoh and Nor Aini Abdul Rahman

The Effects of Pre-Composting Subjected to Black Soldier Fly Larvae Digestion on Agri-Food Waste Maturation Indicators 2521
Muhammad Heikal Ismail, Naweedullah Amin, Anie Yulistyorini, Ridwan Muhamad Rifai, Khursheed Alam Khan and Luqman Chuah Abdullah

State-of-the-Art CNN Models for Plant Disease Classification: A Comparative Study 2541
Tanko Daniel Salka, Marsyita Hanafi, Sharifah M. Syed Ahmad Abdul Rahman and Dzarifah Mohamed Zulperi

Pertanika Editorial Office, Journal Division,
Putra Science Park,
1st Floor, IDEA Tower II,
UPM-MTDC Center,
Universiti Putra Malaysia,
43400 UPM Serdang,
Selangor Darul Ehsan
Malaysia

http://www.pertanika.upm.edu.my
Email: executive_editor.pertanika@upm.edu.my
Tel. No.: +603- 9769 1622

PENERBIT UPM UNIVERSITI PUTRA MALAYSIA PRESS

http://penerbit.upm.edu.my
Email: penerbit@upm.edu.my
Tel. No.: +603- 9769 8855

e-ISSN 2231-8526